Chapter 8
The Minimax Principle

8.1 Tests with Guaranteed Power

The criteria discussed so far, unbiasedness and invariance, suffer from the disadvantage of being applicable, or leading to optimum solutions, only in rather restricted classes of problems. We shall therefore turn now to an alternative approach, which potentially is of much wider applicability. Unfortunately, its application to specific problems is in general not easy, unless there exists a UMP invariant test.

One of the important considerations in planning an experiment is the number of observations required to ensure that the resulting statistical procedure will have the desired precision or sensitivity. For problems of hypothesis testing this means that the probabilities of the two kinds of errors should not exceed certain preassigned bounds, say α and $1-\beta$, so that the tests must satisfy the conditions

$$\begin{aligned} E_\theta\varphi(X) &\le \alpha \quad \text{for}\,\theta \in \Omega_H, \\ E_\theta\varphi(X) &\ge \beta \quad \text{for}\,\theta \in \Omega_K. \end{aligned} \tag{8.1}$$

If the power function $E_\theta\varphi(X)$ is continuous and if $\alpha < \beta$, (8.1) cannot hold when the sets Ω_H and Ω_K are contiguous. This mathematical difficulty corresponds in part to the fact that the division of the parameter values θ into the classes Ω_H and Ω_K for which the two different decisions are appropriate is frequently not sharp. Between the values for which one or the other of the decisions is clearly correct there may lie others for which the relative advantages and disadvantages of acceptance and rejection are approximately in balance. Accordingly we shall assume that Ω is partitioned into three sets

$$\Omega = \Omega_H + \Omega_I + \Omega_K,$$

of which Ω_I designates the *indifference zone*, and Ω_K the class of parameter values differing so widely from those postulated by the hypothesis that false acceptance of H is a serious error, which should occur with probability at most $1-\beta$.

E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, Springer Texts in Statistics, https://doi.org/10.1007/978-3-030-70578-7_8

To see how the sample size is determined in this situation, suppose that X_1, $X_2, \ldots$ constitute the sequence of available random variables, and for a moment let n be fixed and let $X = (X_1, \ldots, X_n)$. In the usual applications (for a more precise statement, see Problem 8.1), there exists a test φ_n which maximizes

$$\inf_{\Omega_k} E_\theta \varphi(X) \tag{8.2}$$

among all level-α tests based on X. Let $\beta_n = \inf_{\Omega_K} E_\theta \varphi_n(X)$, and suppose that for sufficiently large n there exists a test satisfying (8.1). [Conditions under which this is the case are given by Berger (1951a) and Kraft (1955).] The desired sample size, which is the smallest value of n for which $\beta_n \geq \beta$, is then obtained by trial and error. This requires the ability of determining for each fixed n the test that maximizes (8.2) subject to

$$E_\theta \varphi(X) \leq \alpha \quad \text{for} \quad \theta \in \Omega_H. \tag{8.3}$$

A method for determining a test with this *maximin* property (of maximizing the minimum power over Ω_K) is obtained by generalizing Theorem 3.8.1. It will be convenient in this discussion to make a change of notation, and to denote by ω and ω' the subsets of Ω previously denoted by Ω_H and Ω_K. Let $\mathcal{P} = \{P_\theta, \theta \in \omega \cup \omega'\}$ be a family of probability distributions over a sample space $(\mathcal{X}, \mathcal{A})$ with densities $p_\theta = dP_\theta/d\mu$ with respect to a σ-finite measure μ, and suppose that the densities $p_\theta(x)$ considered as functions of the two variables (x, θ) are measurable $(\mathcal{A} \times \mathcal{B})$ and $(\mathcal{A} \times \mathcal{B}')$, where $\mathcal{B}$ and $\mathcal{B}'$ are given σ-fields over ω and ω'. Under these assumptions, the following theorem gives conditions under which a solution of a suitable Bayes problem provides a test with the required properties.

Theorem 8.1.1 *For any distributions Λ and Λ' over $\mathcal{B}$ and $\mathcal{B}'$, let $\varphi_{\Lambda,\Lambda'}$ be the most powerful test for testing*

$$h(x) = \int_\omega p_\theta(x)\, d\Lambda(\theta)$$

at level α against

$$h'(x) = \int_{\omega'} p_\theta(x)\, d\Lambda'(\theta)$$

and let $\beta_{\Lambda,\Lambda'}$ be its power against the alternative h'. If there exist Λ and Λ' such that

$$\begin{aligned} \sup_\omega E_\theta \varphi_{\Lambda,\Lambda'}(X) &\leq \alpha, \\ \inf_{\omega'} E_\theta \varphi_{\Lambda,\Lambda'}(X) &= \beta_{\Lambda,\Lambda'}, \end{aligned} \tag{8.4}$$

then:

(i) $\varphi_{\Lambda,\Lambda'}$ *maximizes* $\inf_{\omega'} E_\theta\varphi(X)$ *among all level-*α *tests of the hypothesis* H : $\theta \in \omega$ *and is the unique test with this property if it is the unique most powerful level-*α *test for testing* h *against* h'.

(ii) *The pair of distributions* Λ, Λ' *is least favorable in the sense that for any other pair* ν, ν' *we have*

$$\beta_{\Lambda,\Lambda'} \le \beta_{\nu,\nu'}.$$

PROOF. (i): If φ^* is any other level-α test of H, it is also of level α for testing the simple hypothesis that the density of X is h, and the power of φ^* against h' therefore cannot exceed $\beta_{\Lambda,\Lambda'}$. It follows that

$$\inf_{\omega'} E_\theta\varphi^*(X) \le \int_{\omega'} E_\theta\varphi^*(X)\,d\Lambda'(\theta) \le \beta_{\Lambda,\Lambda'} = \inf_{\omega'} E_\theta\varphi_{\Lambda\Lambda'}(X),$$

and the second inequality is strict if $\varphi_{\Lambda\Lambda'}$ is unique.

(ii): Let ν, ν' be any other distributions over $(\omega, \mathcal{B})$ and $(\omega', \mathcal{B}')$, and let

$$g(x) = \int_\omega p_\theta(x)d\nu(\theta), \qquad g'(x) = \int_{\omega'} p_\theta(x)\,d\nu'(\theta).$$

Since both $\varphi_{\Lambda,\Lambda'}$ and $\varphi_{\nu,\nu'}$ are level-α tests of the hypothesis that $g(x)$ is the density of X, it follows that

$$\beta_{\nu,\nu'} \ge \int \varphi_{\Lambda,\Lambda'}(x)g'(x)\,d\mu(x) \ge \inf_{\omega'} E_\theta\varphi_{\Lambda,\Lambda'}(X) = \beta_{\Lambda,\Lambda'}. \blacksquare$$

Corollary 8.1.1 *Let* Λ, Λ' *be two probability distributions and* C *a constant such that*

$$\varphi_{\Lambda,\Lambda'}(x) = \begin{cases} 1 \text{ if } \int_{\omega'} p_\theta(x)\,d\Lambda'(\theta) > C\int_\omega p_\theta(x)\,d\Lambda(\theta) \\ \gamma \text{ if } \int_{\omega'} p_\theta(x)\,d\Lambda'(\theta) = C\int_\omega p_\theta(x)\,d\Lambda(\theta) \\ 0 \text{ if } \int_{\omega'} p_\theta(x)\,d\Lambda'(\theta) < C\int_\omega p_\theta(x)\,d\Lambda(\theta) \end{cases} \tag{8.5}$$

is a size-α *test for testing that the density of* X *is* $\int_\omega p_\theta(x)\,d\Lambda(\theta)$ *and such that*

$$\Lambda(\omega_0) = \Lambda'(\omega_0') = 1, \tag{8.6}$$

where

$$\omega_0 = \left\{\theta : \theta \in \omega \text{ and } E_\theta\varphi_{\Lambda,\Lambda'}(X) = \sup_{\theta'\in\omega} E_{\theta'}\varphi_{\Lambda,\Lambda'}(X)\right\}$$

$$\omega_0' = \left\{\theta : \theta \in \omega' \text{ and } E_\theta\varphi_{\Lambda,\Lambda'}(X) = \inf_{\theta'\in\omega'} E_{\theta'}\varphi_{\Lambda,\Lambda'}(X)\right\}.$$

Then the conclusions of Theorem 8.1.1 hold.

PROOF. If h, h', and $\beta_{\Lambda,\Lambda'}$ are defined as in Theorem 8.1.1, the assumptions imply that $\varphi_{\Lambda,\Lambda'}$ is a most powerful level-α test for testing h against h', that

$$\sup_{\omega} E_\theta \varphi_{\Lambda,\Lambda'}(X) = \int_{\omega} E_\theta \varphi_{\Lambda,\Lambda'}(X)\, d\Lambda(\theta) = \alpha,$$

and that

$$\inf_{\omega'} E_\theta \varphi_{\Lambda,\Lambda'}(X) = \int_{\omega'} E_\theta \varphi_{\Lambda,\Lambda'}(X)\, d\Lambda'(\theta) = \beta_{\Lambda,\Lambda'}.$$

Condition (8.4) is thus satisfied and Theorem 8.1.1 applies. ■

The following remark is often useful in applying the theorem. Suppose $\varphi_{\Lambda,\Lambda'}$ satisfies: its power function is constant and smallest over ω' on $\omega'_0 \equiv$ the support of Λ'. Then, the condition

$$\beta_{\Lambda,\Lambda'} = \inf_{\theta \in \omega'} E_\theta \varphi_{\Lambda,\Lambda'}$$

holds. To see why, note that

$$\inf_{\theta \in \omega'} E_\theta \varphi_{\Lambda,\Lambda'} = \inf_{\theta \in \omega'_0} E_\theta \varphi_{\Lambda,\Lambda'} = \int_{\theta \in \omega'} E_\theta \varphi_{\Lambda,\Lambda'} d\Lambda'(\theta)$$

$$= \int \int \varphi_{\Lambda,\Lambda'} p_\theta(x) d\Lambda'(\theta) \mu(dx) = \int \varphi_{\Lambda,\Lambda'} h'(x) \mu(dx) = \beta_{\Lambda,\Lambda'}\ .$$

Example 8.1.1 (**Simple example**) Suppose $X \sim N(\xi, 1)$. Test $H : \xi = 0$ versus $H' : |\xi| \geq \epsilon$, where $\epsilon > 0$ is fixed. Let Λ' put equal mass at $\pm\epsilon$. To see why this works, calculate the Neyman–Pearson test $\varphi_{\Lambda,\Lambda'}$ given (8.5); it rejects for large values of

$$\frac{\exp[-\frac{1}{2}(X-\epsilon)^2] + \exp[-\frac{1}{2}(X+\epsilon)^2]}{\exp(-\frac{1}{2}X^2)} \propto [\exp(\epsilon X) + \exp(-\epsilon X)]\ .$$

The last expression is clearly a function of $|X|$ and it is easy to check that it is an increasing function of $|X|$. So, the test rejects for large $|X|$, i.e., when $|X| \geq z_{1-\frac{\alpha}{2}}$. Its minimum power occurs when $|\xi| = \epsilon$ (because the family of distributions of $|X|$ has monotone likelihood ratio), and its power is the same at ϵ and $-\epsilon$. So, it is maximin. It is also UMPI. However, it is not UMPU for the alternatives considered (Problem 8.2). ■

Example 8.1.2 (**Many normal means**) Suppose $X_1, \ldots, X_n$ are independent with $X_i \sim N(\xi_i, 1)$. The null hypothesis specifes $H : \xi_1 = \cdots \xi_n = 0$ while the alternative specifies exactly one of the $\xi_i = \xi$, where $\xi > 0$ is assumed known. The least favorable distribution Λ' is uniform on the n vectors

$$(\xi, 0, \ldots, 0),\ (0, \xi, 0, \ldots, 0), \ldots, (0, \ldots, 0, \xi)\ .$$

The resulting test $\varphi_{\Lambda,\Lambda'}$ rejects for large values of $T = \sum_{i=1}^{n} \exp(\xi X_i)$ To see why, the average likelihood under the alternative divided by the likelihood under the null is given by:

$$\frac{\frac{1}{n}\sum_{i=1}^{n} \exp[-\frac{1}{2}(X_i - \xi)^2]\Pi_{j\neq i} \exp[-\frac{1}{2}X_j^2]}{\Pi_{i=1}^{n} \exp[-\frac{1}{2}X_i^2]} ,$$

which is equivalent to T. By symmetry, its power is the same for all n alternative mean vectors, and so it is maximin. ■

Example 8.1.3 **(Many normal means with different** ω'**)** Under the setup of Example 8.1.2, suppose the alternative parameter space is specified by

$$\omega' = \{(\xi_1, \ldots, \xi_n) : \sum_i \xi_i^2 = \delta^2\} ,$$

where $\delta > 0$ is fixed. Now, let Λ' be the uniform distribution on S_n, the sphere of radius δ centered at the origin. Letting U denote the uniform distribution on S_n, the test $\varphi_{\Lambda,\Lambda'}$ rejects for large values of

$$\frac{\int_{S_n} \Pi_i \exp[-\frac{1}{2}(X_i - \xi_i)^2] dU(\xi)}{\Pi_i \exp[-\frac{1}{2}X_i^2]}$$

$$\propto \int_{S_n} \exp(\sum_i \xi_i X_i) dU(\xi) = C E[\exp(\Xi^\top X)] ,$$

where $\Xi = (\Xi_1, \ldots, \Xi_n)^\top$ is random and uniform on S_n and the above expectation is with respect to the distribution of the Ξ_i with the X_i fixed.

By symmetry, the test $\varphi_{\Lambda,\Lambda'}$ must be invariant with respect to orthogonal transformations. Indeed, $E[\exp(\Xi^\top X)]$ is the same for $X = (X_1, \ldots, X_n)^\top$ and OX, if O is orthogonal. To see why, note that $\Xi^\top OX = (O\Xi)^\top X$. Since the distribution of $O\Xi$ and Ξ are the same, then the distribution of $\Xi^\top OX$ is the same as that of $\Xi^\top X$, giving the result.

Therefore, the maximin test is invariant, and so it cannot improve upon the UMPI test. Therefore, the UMPI test must be maximin. By Example 6.3.3, the UMPI test rejects for large values of $T = \sum_i X_i^2$. If we did not know the UMPI test already, we just need to show that $E[\exp(\Xi^\top X)]$ is an increasing function of T; see Problem 8.3. ■

Suppose that the sets Ω_H, Ω_I, and Ω_K are defined in terms of a nonnegative function d, which is a measure of the distance of θ from H, by

$$\Omega_H = \{\theta : d(\theta) = 0\}, \qquad \Omega_I = \{\theta : 0 < d(\theta) < \Delta\},$$
$$\Omega_K = \{0 : d(\theta) \geq \Delta\}.$$

Suppose also that the power function of any test is continuous in θ. In the limit as $\Delta = 0$, there is no indifference zone. Then Ω_K becomes the set $\{\theta : d(\theta) > 0\}$, and the infimum of $\beta(\theta)$ over Ω_K is $\leq \alpha$ for any level-α test. This infimum is therefore maximized by any test satisfying $\beta(\theta) \geq \alpha$ for all $\theta \in \Omega_K$, that is, by any unbiased test, so that unbiasedness is seen to be a limiting form of the maximin criterion. A more useful limiting form, since it will typically lead to a unique test, is given by the following definition. A test φ_0 is said to *maximize the minimum power locally*[1] if, given any other test φ, there exists Δ_0 such that

$$\inf_{\omega_\Delta} \beta_{\varphi_0}(\theta) \geq \inf_{\omega_\Delta} \beta_{\varphi}(\theta) \qquad \text{for all} \quad 0 < \Delta < \Delta_0, \tag{8.7}$$

where ω_Δ is the set of θ's for which $d(\theta) \geq \Delta$.

8.2 Further Examples

In Chapter 3 it was shown for a family of probability densities depending on a real parameter θ that a UMP test exists for testing $H : \theta \leq \theta_0$ against $\theta > \theta_0$ provided for all $\theta < \theta'$ the ratio $p_{\theta'}(x)/p_\theta(x)$ is a monotone function of some real-valued statistic. This assumption, although satisfied for a one-parameter exponential family, is quite restrictive, and a UMP test of H will in fact exist only rarely. A more general approach is furnished by the formulation of the preceding section. If the indifference zone is the set of θ's with $\theta_0 < \theta < \theta_1$, the problem becomes that of maximizing the minimum power over the class of alternatives $\omega' : \theta \geq \theta_1$. Under appropriate assumptions, one would expect the least favorable distributions Λ and Λ' of Theorem 8.1.1 to assign probability 1 to the points θ_0 and θ_1, and hence the maximin test to be given by the rejection region $p_{\theta_1}(x)/p_{\theta_0}(x) > C$. The following lemma gives sufficient conditions for this to be the case.

Lemma 8.2.1 *Let $X_1, \ldots, X_n$ be identically and independently distributed with probability density $f_\theta(x)$, where θ and x are real-valued, and suppose that for any $\theta < \theta'$ the ratio $f_{\theta'}(x)/f_\theta(x)$ is a nondecreasing function of x. Then the level-α test φ of H which maximizes the minimum power over ω' is given by*

$$\varphi(x_1, \ldots, x_1) = \begin{cases} 1 \text{ if } & r(x_1, \ldots, x_n) > C, \\ \gamma \text{ if } & r(x_1, \ldots, x_n) = C, \\ 0 \text{ if } & r(x_1, \ldots, x_n) < C, \end{cases} \tag{8.8}$$

where $r(x_1, \ldots, x_n) = f_{\theta_1}(x_1) \ldots f_{\theta_1}(x_n)/f_{\theta_0}(x_1) \ldots f_{\theta_0}(x_n)$ and where C and γ are determined by

$$E_{\theta_0}\varphi(X_1, \ldots, X_n) = \alpha. \tag{8.9}$$

[1] A different definition of local minimaxity is given by Giri and Kiefer (1964).

PROOF. The function $\varphi(x_1, \dots, x_n)$ is nondecreasing in each of its arguments, so that by Lemma 3.4.2,

$$E_\theta \varphi(X_1, \dots, X_n) \leq E_{\theta'} \varphi(X_1, \dots, X_n)$$

when $\theta < \theta'$. Hence the power function of φ is monotone and φ is a level-α test. Since $\varphi = \varphi_{\Lambda,\Lambda'}$, where Λ and Λ' are the distributions assigning probability 1 to the points θ_0 and θ_1, Condition (8.4) is satisfied, which proves the desired result as well as the fact that the pair of distributions (Λ, Λ') is least favorable. ■

Example 8.2.1 Let θ be a location parameter, so that $f_\theta(x) = g(x - \theta)$, and suppose for simplicity that $g(x) > 0$ for all x. We will show that a necessary and sufficient condition for $f_\theta(x)$ to have monotone likelihood ratio in x is that $-\log g$ is convex. The condition of monotone likelihood ratio in x,

$$\frac{g(x - \theta')}{g(x - \theta)} \leq \frac{g(x' - \theta')}{g(x' - \theta)} \quad \text{for all} \quad x < x', \quad \theta < \theta',$$

is equivalent to

$$\log g(x' - \theta) + \log g(x - \theta') \leq \log g(x - \theta) + \log g(x' - \theta').$$

Since $x - \theta = t(x - \theta') + (1 - t)(x' - \theta)$ and $x' - \theta' = (1 - t)(x - \theta') + t(x' - \theta)$, where $t = (x' - x)/(x' - x + \theta' - \theta)$, a sufficient condition for this to hold is that the function $-\log g$ is convex. To see that this condition is also necessary, let $a < b$ be any real numbers, and let $x - \theta' = a$, $x' - \theta = b$, and $x' - \theta' = x - \theta$. Then $x - \theta = \frac{1}{2}(x' - \theta + x - \theta') = \frac{1}{2}(a + b)$, and the condition of monotone likelihood ratio implies

$$\tfrac{1}{2}[\log g(a) + \log g(b)] \leq \log g\left[\tfrac{1}{2}(a + b)\right].$$

Since $\log g$ is measurable, this in turn implies that $-\log g$ is convex.[2]

A density g for which $-\log g$ is convex is called *strongly unimodal*. Basic properties of such densities were obtained by Ibragimov (1956). Strong unimodality is a special case of total positivity. A density of the form $g(x - \theta)$ which is totally positive of order r is said to be a Polya frequency function of order r. It follows from Example 8.2.1 that $g(x - \theta)$ is a Polya frequency function of order 2 if and only if it is strongly unimodal. [For further results concerning Polya frequency functions and strongly unimodal densities, see Karlin (1968), Marshall and Olkin (1979), Huang and Ghosh (1982), and Loh (1984a, b).]

Two distributions which satisfy the above condition [besides the normal distribution, for which the resulting densities $p_\theta(x_1, \dots, x_n)$ form an exponential family] are the *double-exponential distribution* with

$$g(x) = \tfrac{1}{2} e^{-|x|}$$

[2] See Sierpinski (1920).

and the *logistic distribution*, whose cumulative distribution function is

$$G(x) = \frac{1}{1 + e^{-x}},$$

so that the density is $g(x) = e^{-x}/(1 + e^{-x})^2$. ■

Example 8.2.2 To consider the corresponding problem for a scale parameter, let $f_\theta(x) = \theta^{-1}h(x/\theta)$ where h is an even function. Without loss of generality one may then restrict x to be nonnegative, since the absolute values $|X_1|, \ldots, |X_n|$ form a set of sufficient statistics for θ. If $Y_i = \log X_i$ and $\eta = \log \theta$, the density of Y_i is

$$h(e^{y-\eta})e^{y-\eta}.$$

By Example 8.2.1, if $h(x) > 0$ for all $x \geq 0$, a necessary and sufficient condition for $f_{\theta'}(x)/f_\theta(x)$ to be a nondecreasing function of x for all $\theta < \theta'$ is that $-\log[e^y h(e^y)]$ or equivalently $-\log h(e^y)$ is a convex function of y. An example in which this holds—in addition to the normal and double-exponential distributions, where the resulting densities form an exponential family—is the *Cauchy distribution* with

$$h(x) = \frac{1}{\pi}\frac{1}{1 + x^2}.$$

Since the convexity of $-\log h(y)$ implies that of $-\log h(e^y)$, it follows that if h is an even function and $h(x - \theta)$ has monotone likelihood ratio, so does $h(x/\theta)$. When h is the normal or double-exponential distribution, this property of $h(x/\theta)$ also follows from Example 8.2.1. That monotone likelihood ratio for the scale-parameter family does not conversely imply the same property for the associated location parameter family is illustrated by the Cauchy distribution. The condition is therefore more restrictive for a location than for a scale parameter. ■

The chief difficulty in the application of Theorem 8.1.1 to specific problems is the necessity of knowing, or at least being able to guess correctly, a pair of least favorable distributions (Λ, Λ'). Guidance for obtaining these distributions is sometimes provided by invariance considerations. If there exists a group G of transformations of X such that the induced group $\bar{G}$ leaves both ω and ω' invariant, the problem is symmetric in the various θ's that can be transformed into each other under $\bar{G}$. It then seems plausible that unless Λ and Λ' exhibit the same symmetries, they will make the statistician's task easier, and hence will not be least favorable.

Example 8.2.3 In the problem of paired comparisons considered in Example 6.3.6, the observations X_i $(i = 1, \ldots, n)$ are independent variables taking on the values 1 and 0 with probabilities p_i and $q_i = 1 - p_i$. The hypothesis H to be tested specifies the set $\omega : \max p_i \leq \frac{1}{2}$. Only alternatives with $p_i \geq \frac{1}{2}$ for all i are considered, and as ω' we take the subset of those alternatives for which $\max p_i \geq \frac{1}{2} + \delta$. One would expect Λ to assign probability 1 to the point $p_1 = \cdots p_n = \frac{1}{2}$, and Λ' to assign positive probability only to the n points $(p_1, \ldots, p_n)$ which have $n - 1$ coordinates equal to $\frac{1}{2}$ and the remaining coordinate equal to $\frac{1}{2} + \delta$. Because of the symmetry with regard to the n variables, it seems plausible that Λ' should assign equal probability $1/n$ to each of these n points. With these choices, the test $\varphi_{\Lambda,\Lambda'}$ rejects when

$$\sum_{i=1}^{n} \left(\frac{\frac{1}{2} + \delta}{\frac{1}{2}} \right)^{x_i} > C.$$

This is equivalent to $\sum_{i=1}^{n} x_i > C$, which had previously been seen to be UMP invariant for this problem. Since the critical function $\varphi_{\Lambda,\Lambda'}(x_1, \ldots, x_n)$ is nondecreasing in each of its arguments, it follows from Lemma 3.4.2 that $p_i \leq p'_i$ for $i = 1, \ldots, n$ implies

$$E_{p_1,\ldots,p_n} \varphi_{\Lambda,\Lambda'}(X_1, \ldots, X_n) \leq E_{p'_1,\ldots,p'_n} \varphi_{\Lambda,\Lambda'}(X_1, \ldots, X_n)$$

and hence the conditions of Theorem 8.1.1 are satisfied. ■

Example 8.2.4 Let $X = (X_1, \ldots, X_n)$ be a sample from $N(\xi, \sigma^2)$, and consider the problem of testing $H : \sigma = \sigma_0$ against the set of alternatives $\omega' : \sigma \leq \sigma_1$ or $\sigma \geq \sigma_2$ $(\sigma_1 < \sigma_0 < \sigma_2)$. This problem remains invariant under the transformations $X'_i = X_i + c$, which in the parameter space induce the group $\bar{G}$ of transformations $\xi' = \xi + c$, $\sigma' = \sigma$. One would therefore expect the least favorable distribution Λ over the line $\omega : -\infty < \xi < \infty$, $\sigma = \sigma_0$, to be invariant under $\bar{G}$. Such invariance implies that Λ assigns to any interval a measure proportional to the length of the interval. Hence Λ cannot be a probability measure and Theorem 8.1.1 is not directly applicable. The difficulty can be avoided by approximating Λ by a sequence of probability distributions, in the present case, for example, by the sequence of normal distributions $N(0, k)$, $k = 1, 2, \ldots$.

In the particular problem under consideration, it happens that there also exist least favorable distributions Λ and Λ', which are true probability distributions and therefore not invariant. These distributions can be obtained by an examination of the corresponding one-sided problem in Section 3.9, as follows. On ω, where the only variable is ξ, the distribution Λ of ξ is taken as the normal distribution with an arbitrary mean ξ_1 and with variance $(\sigma_2^2 - \sigma_0^2)/n$. Under Λ' all probability should be concentrated on the two lines $\sigma = \sigma_1$ and $\sigma = \sigma_2$ in the (ξ, σ) plane, and we put $\Lambda' = p\Lambda'_1 + q\Lambda'_2$, where Λ'_1 is the normal distribution with mean ξ_1 and variance $(\sigma_2^2 - \sigma_1^2)/n$, while Λ'_2 assigns probability 1 to the point (ξ_1, σ_2). A computation analogous to that carried out in Section 3.9 then shows the acceptance region to be given by

$$\frac{\dfrac{p}{\sigma_1^{n-1}\sigma_2}\exp\left[\dfrac{-1}{2\sigma_1^2}\sum(x_i-\bar{x})^2-\dfrac{n}{2\sigma_2^2}(\bar{x}-\xi_1)^2\right] + \dfrac{q}{\sigma_2^n}\exp\left[\dfrac{-1}{2\sigma_2^2}\left\{\sum(x_i-\bar{x})^2+n(\bar{x}-\xi_1)^2\right\}\right]}{\dfrac{1}{\sigma_0^{n-1}\sigma_2}\exp\left[\dfrac{-1}{2\sigma_0^2}\sum(x_i-\bar{x})^2-\dfrac{n}{2\sigma_2^2}(\bar{x}-\xi_1)^2\right]} < C\,,$$

which is equivalent to

$$C_1 \le \sum(x_i-\bar{x})^2 \le C_2.$$

The probability of this inequality is independent of ξ, and hence C_1 and C_2 can be determined so that the probability of acceptance is $1-\alpha$ when $\sigma=\sigma_0$, and is equal for the two values $\sigma=\sigma_1$ and $\sigma=\sigma_2$.

It follows from Section 3.7 that there exist p and C which lead to these values of C_1 and C_2 and that the above test satisfies the conditions of Corollary 8.1.1 with $\omega_0=\omega$, and with ω_0' consisting of the two lines $\sigma=\sigma_1$ and $\sigma=\sigma_2$. ■

8.3 Comparing Two Approximate Hypotheses

As in Section 3.2, let $P_0 \ne P_1$ be two distributions possessing densities p_0 and p_1 with respect to a measure μ. Since distributions even at best are known only approximately, let us assume that the true distributions are approximately P_0 or P_1 in the sense that they lie in one of the families

$$\mathcal{P}_i = \{Q : Q = (1-\epsilon_i)P_i + \epsilon_i G_i\}, \qquad i = 0, 1, \tag{8.10}$$

with ϵ_0, ϵ_1 given and the G_i arbitrary unknown distributions. We wish to find the level-α test of the hypothesis H that the true distribution lies in $\mathcal{P}_0$, which maximizes the minimum power over $\mathcal{P}_1$. This is the problem considered in Section 8.1 with θ indicating the true distribution, $\Omega_H = \mathcal{P}_0$, and $\Omega_K = \mathcal{P}_1$.

The following theorem shows the existence of a pair of least favorable distributions Λ and Λ' satisfying the conditions of Theorem 8.1.1, each assigning probability 1 to a single distribution, Λ to $Q_0 \in \mathcal{P}_0$ and Λ' to $Q_1 \in \mathcal{P}_1$, and exhibits the Q_i explicitly.

Theorem 8.3.1 *Let*

$$q_0(x) = \begin{cases} (1-\epsilon_0)p_0(x) & \text{if } \frac{p_1(x)}{p_0(x)} < b, \\ \frac{(1-\epsilon_0)p_1(x)}{b} & \text{if } \frac{p_1(x)}{p_0(x)} \ge b, \end{cases}$$
$$q_1(x) = \begin{cases} (1-\epsilon_1)p_1(x) & \text{if } \frac{p_1(x)}{p_0(x)} > a, \\ a(1-\epsilon_1)p_0(x) & \text{if } \frac{p_1(x)}{p_0(x)} \le a. \end{cases} \tag{8.11}$$

(i) *For all* $0 < \epsilon_i < 1$, *there exist unique constants* a *and* b *such that* q_0 *and* q_1 *are probability densities with respect to* μ; *the resulting* q_i *are members of* $\mathcal{P}_i$ $(i = 0, 1)$.
(ii) *There exist* δ_0, δ_1 *such that for all* $\epsilon_i \le \delta_i$ *the constants* a *and* b *satisfy* $a < b$ *and that the resulting* q_0 *and* q_1 *are distinct.*
(iii) *If* $\epsilon_i \le \delta_i$ *for* $i = 0, 1$, *the families* $\mathcal{P}_0$ *and* $\mathcal{P}_1$ *are nonoverlapping and the pair* (q_0, q_1) *is least favorable, so that the maximin test of* $\mathcal{P}_0$ *against* $\mathcal{P}_1$ *rejects when* $q_1(x)/q_0(x)$ *is sufficiently large.*

Note. Suppose $a < b$, *and let*

$$r(x) = \frac{p_1(x)}{p_0(x)}, \quad r^*(x) = \frac{q_1(x)}{q_0(x)}, \quad \textit{and} \quad k = \frac{1-\epsilon_1}{1-\epsilon_0}.$$

Then

$$r^*(x) = \begin{cases} ka & \textit{when} \quad r(x) \le a, \\ kr(x) & \textit{when} \quad a < r(x) < b, \\ kb & \textit{when} \quad b \le r(x). \end{cases} \tag{8.12}$$

The maximin test thus replaces the original probability ratio with a censored version.

PROOF. The proof will be given under the simplifying assumption that $p_0(x)$ and $p_1(x)$ are positive for all x in the sample space.

(i): For q_1 to be a probability density, a must satisfy the equation

$$P_1[r(X) > a] + aP_0[r(X) \le a] = \frac{1}{1-\epsilon_1}. \tag{8.13}$$

If (8.13) holds, it is easily checked that $q_1 \in \mathcal{P}_1$ (Problem 8.15). To prove existence and uniqueness of a solution a of (8.13), let

$$\gamma(c) = P_1[r(X) > c] + cP_0[r(X) \le c].$$

Then

$$\gamma(0) = 1 \quad \text{and} \quad \gamma(c) \to \infty \quad \text{as } c \to \infty. \tag{8.14}$$

Furthermore (Problem 8.17)

$$\gamma(c+\Delta) - \gamma(c) = \Delta \int_{r(x) \le c} p_0(x)\, d\mu(x) + \int_{c < r(x) \le c+\Delta} [c + \Delta - r(x)] p_0(x)\, d\mu(x). \tag{8.15}$$

It follows from (8.15) that $0 \le \gamma(c+\Delta) - \gamma(c) \le \Delta$, so that $-\gamma$ is continuous and nondecreasing. Together with (8.14) this establishes the existence of a solution. To prove uniqueness, note that

$$\gamma(c + \Delta) - \gamma(c) \geq \Delta \int_{r(x)<c} p_0(x)\, d\mu(x) \tag{8.16}$$

and that $\gamma(c) = 1$ for all c for which

$$P_i[r(x) \leq c] = 0 \qquad (i = 0, 1). \tag{8.17}$$

If c_0 is the supremum of the values for which (8.17) holds, (8.16) shows that γ is strictly increasing for $c > c_0$ and this proves uniqueness. The proof for b is exactly analogous (Problem 8.16).

(ii): As $\epsilon_1 \to 0$, the solution a of (8.13) tends to c_0. Analogously, as $\epsilon_1 \to 0$, $b \to \infty$ (Problem 8.16).

(iii): This will follow from the following facts:

(a) When X is distributed according to a distribution in $\mathcal{P}_0$, the statistic $r^*(X)$ is stochastically largest when the distribution of X is Q_0.
(b) When X is distributed according to a distribution in $\mathcal{P}_1$, $r^*(X)$ is stochastically smallest for Q_1.
(c) $r^*(X)$ is stochastically larger when the distribution of X is Q_1 than when it is Q_0.

These statements are summarized in the inequalities

$$Q_0'[r^*(X) < t] \geq Q_0[r^*(X) < t] \geq Q_1[r^*(X) < t] \geq Q_1'[r^*(X) < t] \tag{8.18}$$

for all t and all $Q_i' \in \mathcal{P}_i$.

From (8.12), it is seen that (8.18) is obvious when $t \leq ka$ or $t > kb$. Suppose therefore that $ak < t \leq bk$, and denote the event $r^*(X) < t$ by E. Then $Q_0'(E) \geq (1 - \epsilon_0)P_0(E)$ by (8.10). But $r^*(x) < t < kb$ implies $r(X) < b$ and hence $Q_0(E) = (1 - \epsilon)P_0(E)$. Thus $Q_0'(E) \geq Q_0(E)$, and analogously $Q_1'(E) \leq Q_1(E)$. Finally, the middle inequality of (8.18) follows from Corollary 3.2.1.

If the ϵ's are sufficiently small so that $Q_0 \neq Q_1$, it follows from (a)–(c) that $\mathcal{P}_0$ and $\mathcal{P}_1$ are nonoverlapping.

That (Q_0, Q_1) is least favorable and the associated test φ is maximin now follows from Theorem 8.1.1, since the most powerful test φ for testing Q_0 against Q_1 is a nondecreasing function of $q_1(X)/q_0(X)$. This shows that $E\varphi(X)$ takes on its sup over $\mathcal{P}_0$ at Q_0 and its inf over $\mathcal{P}_1$ at Q_1, and this completes the proof. ■

Generalizations of this theorem are given by Huber and Strassen (1973,1974). See also Rieder (1977) and Bednarski (1984). An optimum permutation test, with generalizations to the case of unknown location and scale parameters, is discussed by Lambert (1985).

When the data consist of n identically, independently distributed random variables $X_1, \ldots, X_n$, the neighborhoods (8.10) may not be appropriate, since they do not preserve the assumption of independence. If P_i has density

$$p_i(x_1, \ldots, x_n) = f_i(x_1) \ldots f_i(x_n) \qquad (i = 0, 1), \tag{8.19}$$

a more appropriate model approximating (8.19) may then assign to $X = (X_1, \ldots, X_n)$ the family $\mathcal{P}_i^*$ of distributions according to which the X_j are independently distributed, each with distribution

$$(1-\epsilon_i)F_i(x_j) + \epsilon_i G_i(x_j), \tag{8.20}$$

where F_i has density f_i and where as before the G_i are arbitrary.

Corollary 8.3.1 *Suppose q_0 and q_1 defined by (8.11) with $x = x_j$ satisfy (8.18) and hence are a least favorable pair for testing $\mathcal{P}_0$ against $\mathcal{P}_1$ on the basis of the single observation X_j. Then the pair of distributions with densities $q_i(x_1)\ldots q_i(x_n)$ ($i = 0, 1$) is least favorable for testing $\mathcal{P}_0^*$ against $\mathcal{P}_1^*$, so that the maximin test is given by*

$$\varphi(x_1, \ldots, x_n) = \begin{cases} 1 \\ \gamma \quad \text{if} \prod_{j=1}^{n} \left[\dfrac{q_1(x_j)}{q_0(x_j)}\right] \gtreqless c. \\ 0 \end{cases} \tag{8.21}$$

Proof. By assumption, the random variables $Y_j = q_1(X_j)/q_0(X_j)$ are stochastically increasing as one moves successively from $Q_0' \in \mathcal{P}_0$ to Q_0 to Q_1 to $Q_1' \in \mathcal{P}_1$. The same is then true of any function $\psi(Y_1, \ldots, Y_n)$ which is nondecreasing in each of its arguments by Lemma 3.4.1, and hence of φ defined by (8.21). The proof now follows from Theorem 8.3.1. ■

Instead of the problem of testing P_0 against P_1, consider now the situation of Lemma 8.2.1 where $H : \theta \leq \theta_0$ is to be tested against $\theta \geq \theta_1$ ($\theta_0 < \theta_1$) on the basis of n independent observations X_j, each distributed according to a distribution $F_\theta(x_j)$ whose density $f_\theta(x_j)$ is assumed to have monotone likelihood ratio in x_j.

A robust version of this problem is obtained by replacing F_θ with

$$(1-\epsilon)F_\theta(x_j) + \epsilon G(x_j), \qquad j = 1, \ldots, n, \tag{8.22}$$

where ϵ is given and for each θ the distribution G is arbitrary. Let $\mathcal{P}_0^{**}$ and $\mathcal{P}_1^{**}$ be the classes of distributions (8.22) with $\theta \leq \theta_0$ and $\theta \geq \theta_1$, respectively; and let $\mathcal{P}_0^*$ and $\mathcal{P}_1^*$ be defined as in Corollary 8.3.1 with f_{θ_i} in place of f_i. Then the maximin test (8.21) of $\mathcal{P}_0^*$ against $\mathcal{P}_1^*$ retains this property for testing $\mathcal{P}_0^{**}$ against $\mathcal{P}_1^{**}$.

This is proved in the same way as Corollary 8.3.1, using the additional fact that if $F_{\theta'}$ is stochastically larger than F_θ, then $(1-\epsilon)F_{\theta'} + \epsilon G$ is stochastically larger than $(1-\epsilon)F_\theta + \epsilon G$.

8.4 Maximin Tests and Invariance

When the problem of testing Ω_H against Ω_K remains invariant under a certain group of transformations, it seems reasonable to expect the existence of an invariant pair of least favorable distributions (or at least of sequences of distributions which in some sense are least favorable and invariant in the limit), and hence also of a maximin test which is invariant. This suggests the possibility of bypassing the somewhat cumbersome approach of the preceding sections. If it could be proved that for an invariant problem there always exists an invariant test that maximizes the minimum power over Ω_K, attention could be restricted to invariant tests; in particular, a UMP invariant test would then automatically have the desired maximin property (although it would not necessarily be admissible). These speculations turn out to be correct for an important class of problems, although unfortunately not in general. To find out under what conditions they hold, it is convenient first to separate out the statistical aspects of the problem from the group-theoretic ones by means of the following lemma.

Lemma 8.4.1 *Let $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$ be a dominated family of distributions on $(\mathcal{X}, \mathcal{A})$, and let G be a group of transformations of $(\mathcal{X}, \mathcal{A})$, such that the induced group $\bar{G}$ leaves the two subsets Ω_H and Ω_K of Ω invariant. Suppose that for any critical function φ there exists an (almost) invariant critical function ψ satisfying*

$$\inf_{\bar{G}} E_{\bar{g}\theta}\varphi(X) \leq E_\theta \psi(X) \leq \sup_{\bar{G}} E_{\bar{g}\theta}\varphi(X) \tag{8.23}$$

for all $\theta \in \Omega$. Then if there exists a level-α test φ_0 maximizing $\inf_{\Omega_k} E_\theta\varphi(X)$, there also exists an (almost) invariant test with this property.

PROOF. Let $\inf_{\Omega_K} E_\theta\varphi_0(X) = \beta$, and let ψ_0 be an (almost) invariant test such that (8.23) holds with $\varphi = \varphi_0$, $\psi = \psi_0$. Then

$$E_\theta\psi_0(X) \leq \sup_{\bar{G}} E_{\bar{g}\theta}\varphi_0(X) \leq \alpha \qquad \text{for all} \quad \theta \in \Omega_H$$

and

$$E_\theta\psi_0(X) \geq \inf_{\bar{G}} E_{\bar{g}\theta}\varphi_0(X) \geq \beta \qquad \text{for all} \quad \theta \in \Omega_K,$$

as was to be proved. ■

To determine conditions under which there exists an invariant or almost invariant test ψ satisfying (8.23), consider first the simplest case that G is a finite group, $G = \{g_1, \ldots, g_N\}$ say. If ψ is then defined by

$$\psi(x) = \frac{1}{N}\sum_{i=1}^{N} \varphi(g_i x), \tag{8.24}$$

it is clear that ψ is again a critical function, and that it is invariant under G. It also satisfies (8.23), since $E_\theta\varphi(gX) = E_{\bar{g}\theta}\varphi(X)$ so that $E_\theta\psi(X)$ is the average of a number of terms of which the first and last member of (8.23) are the minimum and maximum, respectively.

An illustration of the finite case is furnished by Example 8.2.3. Here the problem remains invariant under the $n!$ permutations of the variables $(X_1, \ldots, X_n)$.

Lemma 8.4.1 is applicable and shows that there exists an invariant test maximizing $\inf_{\Omega_K} E_\theta \varphi(X)$. Thus in particular the UMP invariant test obtained in Example 6.3.6 has this maximin property and therefore constitutes a solution of the problem.

It also follows that, under the setting of Theorem 6.3.1, the UMPI test given by (6.10) is maximin.

The definition (8.24) suggests the possibility of obtaining $\psi(x)$ also in other cases by averaging the values of $\varphi(gx)$ with respect to a suitable probability distribution over the group G. To see what conditions would be required of this distribution, let $\mathcal{B}$ be a σ-field of subsets of G and ν a probability distribution over $(G, \mathcal{B})$. Disregarding measurability problems for the moment, let ψ be defined by

$$\psi(x) = \int \varphi(gx)\, d\nu(g). \tag{8.25}$$

Then $0 \le \psi \le 1$, and (8.23) is seen to hold by applying Fubini's Theorem (Theorem 2.2.4) to the integral of ψ with respect to the distribution P_θ. For any $g_0 \in G$,

$$\psi(g_0 x) = \int \varphi(gg_0 x)\, d\nu(g) = \int \varphi(hx)\, d\nu^*(h)\ ,$$

where $h = gg_0$ and where ν^* is the measure defined by

$$\nu^*(B) = \nu(Bg_0^{-1}) \qquad \text{for all} \quad B \in \mathcal{B},$$

into which ν is transformed by the transformation $h = gg_0$. Thus ψ will have the desired invariance property, $\psi(g_0 x) = \psi(x)$ for all $g_0 \in G$, if ν is *right invariant*, that is, if it satisfies

$$\nu(Bg) = \nu(B) \qquad \text{for all} \quad B \in \mathcal{B}, \quad g \in G. \tag{8.26}$$

Such a condition was previously used in (6.18).

The measurability assumptions required for the above argument are: (i) For any $A \in \mathcal{A}$, the set of pairs (x, g) with $gx \in A$ is measurable $(\mathcal{A} \times \mathcal{B})$. This insures that the function ψ defined by (8.25) is again measurable. (ii) For any $B \in \mathcal{B}$, $g \in G$, the set Bg belongs to $\mathcal{B}$.

Example 8.4.1 If G is a finite group with elements $g_1, \dots, g_N$, let $\mathcal{B}$ be the class of all subsets of G and ν the probability measure assigning probability $1/N$ to each of the N elements. Condition (8.26) is then satisfied, and the definition (8.25) of ψ in this case reduces to (8.24). ■

Example 8.4.2 Consider the group G of orthogonal $n \times n$ matrices Γ, with the group product $\Gamma_1 \Gamma_2$ defined as the corresponding matrix product. Each matrix can be interpreted as the point in n^2-dimensional Euclidean space whose coordinates are the n^2 elements of the matrix. The group then defines a subset of this space; the Borel subsets of G will be taken as the σ-field $\mathcal{B}$. To prove the existence of a right invariant probability measure over $(G, \mathcal{B})$, we shall define a random orthogonal matrix whose probability distribution satisfies (8.26) and is therefore the required measure. With any nonsingular matrix $x = (x_{ij})$, associate the orthogonal matrix $y = f(x)$ obtained by applying the following Gram–Schmidt orthogonalization process to the n row vectors $x_i = (x_{i1}, \dots, x_{in})$ of x : y_1 is the unit vector in the direction of x_1; y_2 the unit vector in the plane spanned by x_1 and x_2 which is orthogonal to y_1 and forms an acute angle with x_2; and so on. Let $y = (y_{ij})$ be the matrix whose ith row is y_i.

Suppose now that the variables X_{ij} $(i, j = 1, \dots, n)$ are independently distributed as $N(0, 1)$, let X denote the random matrix (X_{ij}), and let $Y = f(X)$. To show that the distribution of the random orthogonal matrix Y satisfies (8.26), consider any fixed orthogonal matrix Γ and any fixed set $B \in \mathcal{B}$. Then $P\{Y \in B\Gamma\} = P\{Y\Gamma' \in B\}$ and from the definition of f it is seen that $Y\Gamma' = f(X\Gamma')$. Since the n^2 elements of the matrix $X\Gamma'$ have the same joint distribution as those of the matrix X, the matrices $f(X\Gamma')$ and $f(X)$ also have the same distribution, as was to be proved. ■

Examples 8.4.1 and 8.4.2 are sufficient for the applications to be made here. General conditions for the existence of an invariant probability measure, of which these examples are simple special cases, are given in the theory of Haar measure. [This is treated, for example, in the books by Halmos (1974), Loomis (1953), and Nachbin (1965). For a discussion in a statistical setting, see Eaton (1983, 1989), Farrell (1998a), and Wijsman (1990), and for a more elementary treatment Berger (1985a).]

8.5 The Hunt–Stein Theorem

Invariant measures exist (and are essentially unique) for a large class of groups, but unfortunately they are frequently not finite and hence cannot be taken to be probability measures. The situation is similar and related to that of the nonexistence of a least favorable pair of distributions in Theorem 8.1.1. There it is usually possible to overcome the difficulty by considering instead a sequence of distributions which has the desired property in the limit. Analogously we shall now generalize the construction of ψ as an average with respect to a right-invariant probability distribution, by considering a sequence of distributions over G which are approximately right invariant for n sufficiently large.

Let $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$ be a family of distributions over a Euclidean space $(\mathcal{X}, \mathcal{A})$ dominated by a σ-finite measure μ, and let G be a group of transformations of $(\mathcal{X}, \mathcal{A})$ such that the induced group $\bar{G}$ leaves Ω invariant.

Theorem 8.5.1 (Hunt–Stein.) *Let $\mathcal{B}$ be a σ-field of subsets of G such that for any $A \in \mathcal{A}$ the set of pairs (x, g) with $gx \in A$ is in $\mathcal{A} \times \mathcal{B}$ and for any $B \in \mathcal{B}$ and $g \in G$ the set Bg is in $\mathcal{B}$. Suppose that there exists a sequence of probability distributions ν_n over $(G, \mathcal{B})$ which is asymptotically right invariant in the sense that for any $g \in G$, $B \in \mathcal{B}$,*

$$\lim_{n\to\infty} |\nu_n(Bg) - \nu_n(B)| = 0. \tag{8.27}$$

Then given any critical function φ, there exists a critical function ψ which is almost invariant and satisfies (8.23).

PROOF. Let

$$\psi_n(x) = \int \varphi(gx)\, d\nu_n(g),$$

which as before is measurable and between 0 and 1. By the weak compactness theorem (Theorem A.5.1 of the Appendix) there exists a subsequence $\{\psi_{n_i}\}$ and a measurable function ψ between 0 and 1 satisfying

$$\lim_{i\to\infty} \int \psi_{n_i} p\, d\mu = \int \psi p\, d\mu$$

for all μ-integrable functions p, so that in particular

$$\lim_{i\to\infty} E_\theta \psi_{n_i}(X) = E_\theta \psi(X)$$

for all $\theta \in \Omega$. By Fubini's Theorem,

$$E_\theta \psi_{n_i}(X) = \int [E_\theta \varphi(gX)]\, d\nu_{n_i}(g) = \int E_{\bar{g}\theta} \varphi(X)\, d\nu_{n_i}(g)\ ,$$

so that

$$\inf_{\bar{G}} E_{\bar{g}\theta} \varphi(X) \le E_\theta \psi_{n_i}(X) \le \sup_{\bar{G}} E_{\bar{g}\theta} \varphi(X),$$

and ψ satisfies (8.23).

In order to prove that ψ is almost invariant we shall show below that for all x and g,

$$\psi_{n_i}(gx) - \psi_{n_i}(x) \to 0. \tag{8.28}$$

Let $I_A(x)$ denote the indicator function of a set $A \in \mathcal{A}$. Using the fact that $I_{gA}(gx) = I_A(x)$, we see that (8.28) implies

$$\begin{aligned}
\int_A \psi(x)\, dP_\theta(x) &= \lim_{i\to\infty} \int \psi_{n_i}(x) I_A(x)\, dP_\theta(x) \\
&= \lim_{i\to\infty} \int \psi_{n_i}(gX) I_{gA}(gx)\, dP_\theta(x) \\
&= \int \psi(x) I_{gA}(x)\, dP_{\bar{g}\theta}(x) = \int_A \psi(gx)\, dP_\theta(x)\ ,
\end{aligned}$$

and hence $\psi(gx) = \psi(x)$ (a.e. $\mathcal{P}$), as was to be proved.

To prove (8.28), consider any fixed x and any integer m, and let G be partitioned into the mutually exclusive sets

$$B_k = \left\{ h \in G : a_k < \varphi(hx) \le a_k + \frac{1}{m} \right\}, \qquad k = 0, \dots, m,$$

where $a_k = (k-1)/m$. In particular, B_0 is the set $\{h \in G : \varphi(hx) = 0\}$. It is seen from the definition of the sets B_k that

$$\sum_{k=0}^{m} a_k \nu_{n_i}(B_k) \le \sum_{k=0}^{m} \int_{B_k} \varphi(hx)\, d\nu_{n_i}(h) \le \sum_{k=0}^{m} \left(a_k + \frac{1}{m} \right) \nu_{n_i}(B_k)$$
$$\le \sum_{k=0}^{m} a_k \nu_{n_i}(B_k) + \frac{1}{m},$$

and analogously that

$$\left| \sum_{k=0}^{m} \int_{B_k g^{-1}} \varphi(hgx)\, d\nu_{n_i}(h) - \sum_{k=0}^{m} a_k \nu_{n_i}(B_k g^{-1}) \right| \le \frac{1}{m},$$

from which it follows that

$$\psi_{n_i}(gx) - \psi_{n_i}(x) :\le \sum |a_k| \cdot |\nu_{n_i}(B_k g^{-1}) - \nu_{n_i}(B_k)| + \frac{2}{m}.$$

By (8.27) the first term of the right-hand side tends to zero as i tends to infinity, and this completes the proof. ∎

When there exist a right-invariant measure ν over G and a sequence of subsets G_n of G with $G_n \subseteq G_{n+1}$, $\cup G_n = G$, and $\nu(G_n) = c_n < \infty$, it is suggestive to take for the probability measures ν_n of Theorem 8.5.1 the measures ν/c_n truncated on G_n. This leads to the desired result in the example below. On the other hand, there are cases in which there exists such a sequence of subsets of G_n but no invariant test satisfying (8.23) and hence no sequence ν_n satisfying (8.27).

Example 8.5.1 Let $x = (x_1, \dots, x_n)$, $\mathcal{A}$ be the class of Borel sets in n-space, and G the group of translations $(x_1 + g, \dots, x_n + g)$, $-\infty < g < \infty$. The elements of G can be represented by the real numbers, and the group product gg' is then the sum $g + g'$. If $\mathcal{B}$ is the class of Borel sets on the real line, the measurability assumptions of Theorem 8.5.1 are satisfied. Let ν be Lebesgue measure, which is clearly invariant under G, and define ν_n to be the uniform distribution on the interval $I(-n, n) = \{g : -n \le g \le n\}$. Then for all $B \in \mathcal{B}$, $g \in G$,

$$|\nu_n(B) - \nu_n(Bg)| = \frac{1}{2n} |\nu[B \cap I(-n, n)] - \nu[B \cap I(-n-g, n-g)]| \le \frac{|g|}{2n},$$

so that (8.27) is satisfied.

This argument also covers the group of scale transformations $(ax_1, \dots, ax_n)$, $0 < a < \infty$, which can be transformed into the translation group by taking logarithms. ■

When applying the Hunt–Stein Theorem to obtain invariant minimax tests, it is frequently convenient to carry out the calculation in steps, as was done in Theorem 6.6.1. Suppose that the problem remains invariant under two groups D and E, and denote by $y = s(x)$ a maximal invariant with respect to D and by E^* the group defined in Theorem 6.2.2, which E induces in y-space. If D and E^* satisfy the conditions of the Hunt–Stein Theorem, it follows first that there exists a maximin test depending only on $y = s(x)$, and then that there exists a maximin test depending only on a maximal invariant $z = t(y)$ under E^*.

Example 8.5.2 Consider a univariate linear hypothesis in the canonical form in which $Y_1, \dots, Y_n$ are independently distributed as $N(\eta_i, \sigma^2)$, where it is given that $\eta_{s+1} = \cdots = \eta_n = 0$, and where the hypothesis to be tested is $\eta_1 = \cdots = \eta_r = 0$. It was shown in Section 7.1 that this problem remains invariant under certain groups of transformations and that with respect to these groups there exists a UMP invariant test. The groups involved are the group of orthogonal transformations, translation groups of the kind considered in Example 8.5.1, and a group of scale changes. Since each of these satisfies the assumptions of the Hunt–Stein Theorem, and since they leave invariant the problem of maximizing the minimum power over the set of alternatives

$$\sum_{i=1}^{r} \frac{\eta_i^2}{\sigma^2} \geq \psi_1^2 \qquad (\psi_1 > 0), \tag{8.29}$$

it follows that the UMP invariant test of Chapter 7 is also the solution of this maximin problem. It is also seen slightly more generally that the test which is UMP invariant under the same groups for testing

$$\sum_{i=1}^{r} \frac{\eta_i^2}{\sigma^2} \leq \psi_0^2$$

(Problem 7.4) maximizes the minimum power over the alternatives (8.29) for $\psi_0 < \psi_1$. ■

Example 8.5.3 (Stein) Let G be the group of all nonsingular linear transformations of p-space. That for $p > 1$ this does not satisfy the conditions of Theorem 8.5.1 is shown by the following problem, which is invariant under G but for which the UMP invariant test does not maximize the minimum power. Generalizing Example 6.2.1, let $X = (X_1, \dots, X_p)$, $Y = (Y_1, \dots, Y_p)$ be independently distributed according to p-variate normal distributions with zero means and nonsingular covariance matrices $E(X_i X_j) = \sigma_{ij}$ and $E(Y_i Y_j) = \Delta\sigma_{ij}$, and let $H : \Delta \leq \Delta_0$ be tested against $\Delta \geq \Delta_1$ $(\Delta_0 < \Delta_1)$, the σ_{ij} being unknown.

This problem remains invariant if the two vectors are subjected to any common nonsingular transformation, and since with probability 1 this group is transitive over the sample space, the UMP invariant test is trivially $\varphi(x, y) \equiv \alpha$. The maximin power against the alternatives $\Delta \geq \Delta_1$ that can be achieved by invariant tests is therefore α. On the other hand, the test with rejection region $Y_1^2/X_1^2 > C$ has a strictly increasing power function $\beta(\Delta)$, whose minimum over the set of alternatives $\Delta \geq \Delta_1$ is $\beta(\Delta_1) > \beta(\Delta_0) = \alpha$. ■

It is a remarkable feature of Theorem 8.5.1 that its assumptions concern only the group G and not the distributions P_θ.[3] When these assumptions hold for a certain G it follows from (8.23) as in the proof of Lemma 8.4.1 that for any testing problem which remains invariant under G and possesses a UMP invariant test, this test maximizes the minimum power over any invariant class of alternatives. Suppose conversely that a UMP invariant test under G has been shown in a particular problem not to maximize the minimum power, as was the case for the group of linear transformations in Example 8.5.3. Then the assumptions of Theorem 8.5.1 cannot be satisfied. However, this does not rule out the possibility that for another problem remaining invariant under G, the UMP invariant test may maximize the minimum power. Whether or not it does is no longer a property of the group alone but will in general depend also on the particular distributions.

Consider in particular the problem of testing $H : \xi_1 = \cdots = \xi_p = 0$ on the basis of a sample $(X_{\alpha 1}, \ldots, X_{\alpha p})$, $\alpha = 1, \ldots, n$, from a p-variate normal distribution with mean $E(X_{\alpha i}) = \xi_i$ and common covariance matrix $(\sigma_{ij}) = (a_{ij})^{-1}$. This problem remains invariant under a number of groups, including that of all nonsingular linear transformations of p-space, and a UMP invariant test exists. An invariant class of alternatives under these groups is

$$\sum\sum \frac{a_{ij}\xi_i\xi_j}{\sigma^2} \geq \psi_1^2. \tag{8.30}$$

Here, Theorem 8.5.1 is not applicable, and the question of whether the T^2-test of $H : \psi = 0$ maximizes the minimum power over the alternatives

$$\sum\sum a_{ij}\xi_i\xi_j = \psi_1^2 \tag{8.31}$$

[and hence a fortiori over the alternatives (8.30)] presents formidable difficulties. The minimax property was proved for the case $p = 2$, $n = 3$ by Giri, Kiefer, and Stein (1963), for the case $p = 2$, $n = 4$ by Linnik, Pliss, and Salaevskii (1968), and for $p = 2$ and all $n \geq 3$ by Salaevskii (1971). The proof is effected by first reducing the problem through invariance under the group G_1 of Example 6.6.2, to which Theorem 8.5.1 is applicable, and then applying Theorem 8.1.1 to the reduced problem. It is a consequence of this approach that it also establishes the admissibility of T^2 as a test of H against the alternatives (8.31). In view of the inadmissibility

[3] These assumptions are essentially equivalent to the condition that the group G is *amenable*. Amenability and its relationship to the Hunt–Stein Theorem are discussed by Bondar and Milnes (1982) and (with a different terminology) by Stone and von Randow (1968).

results for point estimation when $p \geq 3$ (see Lehmann and Casella (1998), Sections 5.4, 5.5, it seems unlikely that T^2 is admissible for $p \geq 3$, and hence that the same method can be used to prove the minimax property in this situation.

The problem becomes much easier when the minimax property is considered against local or distant alternatives rather than against (8.31). Precise definitions and proofs of the fact that T^2 possesses these properties for all p and n are provided by Giri and Kiefer (1964) and in the references given in Section 7.9.

The theory of this and the preceding section can be extended to confidence sets if the accuracy of a confidence set at level $1 - \alpha$ is assessed by its volume or some other appropriate measure of its size. Suppose that the distribution of X depends on the parameters θ to be estimated and on nuisance parameters ϑ, and that μ is a σ-finite measure over the parameter set $\omega = \{\theta : (\theta, \vartheta) \in \Omega\}$, with ω assumed to be independent of ϑ. Then the confidence sets $S(X)$ for θ are minimax with respect to μ at level $1 - \alpha$ if they minimize

$$\sup E_{\theta,\vartheta}\mu[S(X)]$$

among all confidence sets at the given level.

The problem of minimizing $E\mu[S(X)]$ is related to that of minimizing the probability of covering false values (the criterion for accuracy used so far) by the relation (Problem 8.39)

$$E_{\theta_0,\vartheta}\mu[S(X)] = \int_{\theta \neq \theta_0} P_{\theta_0,\vartheta}[\theta \in S(X)]\, d\mu(\theta), \tag{8.32}$$

which holds provided μ assigns measure zero to the set $\{\theta = \theta_0\}$. (For the special case that θ is real-valued and μ Lebesgue measure, see Problem 5.26.)

Suppose now that the problem of estimating θ is invariant under a group G in the sense of Section 6.11 and that it satisfies the invariance condition

$$\mu[S(gx)] = \mu[S(x)]. \tag{8.33}$$

If uniformly most accurate equivariant confidence sets exist, they minimize (8.32) among all equivariant confidence sets at the given level, and one may hope that under the assumptions of the Hunt–Stein Theorem, they will also be minimax with respect to μ among the class of all (not necessarily equivariant) confidence sets at the given level. Such a result does hold and can be used to show for example that the most accurate equivariant confidence sets of Examples 6.11.2 and 6.11.3 minimize their maximum expected Lebesgue measure. A more general class of examples is provided by the confidence intervals derived from the UMP invariant tests of univariate linear hypotheses such as the confidence spheres for $\theta_i = \mu + \alpha_i$ or for α_i given in Section 7.4.

Minimax confidence sets $S(x)$ are not necessarily admissible; that is, there may exist sets $S'(x)$ having the same confidence level but such that

$$E_{\theta,\vartheta}\mu[S'(X)] \leq E_{\theta,\vartheta}\mu[S(X)] \qquad \text{for all} \quad \theta, \vartheta$$

with strict inequality holding for at least some (θ, ϑ).

Example 8.5.4 Let X_i $(i = 1, \ldots, s)$ be independently normally distributed with mean $E(X_i) = \theta_i$ and variance 1, and let G be the group generated by translations $X_i + c_i$ $(i = 1, \ldots, s)$ and orthogonal transformations of $(X_1, \ldots, X_s)$. (G is the Euclidean group of rigid motions in s-space.) In Example 6.12.2, it was argued that the confidence sets

$$C_0 = \{(\theta_1, \ldots, \theta_s) : \sum(\theta_i - X_i)^2 \leq c\} \tag{8.34}$$

are uniformly most accurate equivariant. The volume $\mu[S(X)]$ of any confidence set $S(X)$ remains invariant under the transformations $g \in G$, and it follows from the results of Problems 8.31 and 8.7 and Examples 8.5.1 and 8.5.2 that the confidence sets (8.34) minimize the maximum expected volume.

However, very surprisingly, they are not admissible unless $s = 1$ or 2. In the case $s \geq 3$, Stein (1962) suggested the region (8.34) can be improved by recentered regions of the form

$$C_1 = \{(\theta_1, \ldots, \theta_s) : (\theta_i - \hat{b}X_i)^2 \leq c\}\ , \tag{8.35}$$

where $\hat{b} = \max(0, 1 - (s-2)/\sum_i X_i^2)$. In fact, Brown (1966) proved that, for $s \geq 3$,

$$P_\theta\{\theta \in C_1\} > P_\theta\{\theta \in C_0\}$$

for all θ. This result, which will not be proved here, is closely related to the inadmissibility of $X_1, \ldots, X_s$ as a point estimator of $(\theta_1, \ldots, \theta_s)$ for a wide variety of loss functions. The work on point estimation, which is discussed in Lehmann and Casella (1998), Sections 5.4–5.6, for squared error loss, provides easier access to these ideas than the present setting. Further entries into the literature on admissibility are Stein (1981), Hwang and Casella (1982), and Tseng and Brown (1997); additional references are provided in Lehmann and Casella (1998), p.423.

The inadmissibility of the confidence sets (8.34) is particularly surprising in that the associated UMP invariant tests of the hypotheses $H : \theta_i = \theta_{i_0}$ $(i = 1, \ldots, s)$ are admissible (Problems 8.29, 8.30). ■

8.6 Most Stringent Tests

One of the practical difficulties in the consideration of tests that maximize the minimum power over a class Ω_K of alternatives is the determination of an appropriate Ω_K. If no information is available on which to base the choice of this set, and if a natural definition is not imposed by invariance arguments, a frequently reasonable definition can be given in terms of the power that can be achieved against the various alternatives. The *envelope power function* β_α^* was defined in Problem 6.27 by

$$\beta_\alpha^*(\theta) = \sup \beta_\varphi(\theta),$$

where β_φ denotes the power of a test φ and where the supremum is taken over all level-α tests of H. Thus $\beta_\alpha^*(\theta)$ is the maximum power that can be attained at level α against the alternative θ. (That it can be attained follows under mild restrictions from Theorem A.5.1 of the Appendix.) If

$$S_\Delta^* = \{\theta : \beta_\alpha^*(\theta) = \Delta\},$$

then of two alternatives $\theta_1 \in S_{\Delta_1}^*$, $\theta_2 \in S_{\Delta_2}^*$, θ_1 can be considered closer to H, equidistant, or further away than θ_2 as Δ_1 is $<$, $=$, or $>$ Δ_2.

The idea of measuring the distance of an alternative from H in terms of the available information has been encountered before. If, for example, $X_1, \ldots, X_n$ is a sample from $N(\xi, \sigma^2)$, the problem of testing $H : \xi \le 0$ was discussed (Section 5.2) both when the alternatives ξ are measured in absolute units and when they are measured in σ-units. The latter possibility corresponds to the present proposal, since it follows from invariance considerations (Problem 6.27) that $\beta_\alpha^*(\xi, \sigma)$ is constant on the lines $\xi/\sigma = \text{constant}$.

Fixing a value of Δ and taking as Ω_K the class of alternatives θ for which $\beta_\alpha^*(\theta) \ge \Delta$, one can determine the test that maximizes the minimum power over Ω_K. Another possibility, which eliminates the need of selecting a value of Δ, is to consider for any test φ the difference $\beta_\alpha^*(\theta) - \beta_\varphi(\theta)$. This difference measures the amount by which the actual power $\beta_\varphi(\theta)$ falls short of the maximum power attainable. A test that minimizes

$$\sup_{\Omega-\omega} [\beta_\alpha^*(\theta) - \beta_\varphi(\theta)] \tag{8.36}$$

is said to be *most stringent*. Thus a test is most stringent if it minimizes its maximum shortcoming.

Let φ_Δ be a test that maximizes the minimum power over S_Δ^*, and hence minimizes the maximum difference between $\beta_\alpha^*(\theta)$ and $\beta_\varphi(\theta)$ over S_Δ^*. If φ_Δ happens to be independent of Δ, it is most stringent. This remark makes it possible to apply the results of the preceding sections to the determination of most stringent tests. Suppose that the problem of testing $H : \theta \in \omega$ against the alternatives $\theta \in \Omega - \omega$ remains invariant under a group G, that there exists a UMP almost invariant test φ_0 with respect to G, and that the assumptions of Theorem 8.5.1 hold. Since $\beta_\alpha^*(\theta)$ and hence the set S^*_Δ is invariant under $\bar{G}$ (Problem 6.27), it follows that φ_0 maximizes the minimum power over S^*_Δ for each Δ, and φ_0 is therefore most stringent.

As an example of this method consider the problem of testing $H : p_1, \dots, p_n \le \frac{1}{2}$ against the alternative $K : p_i > \frac{1}{2}$ for all i, where p_i is the probability of success in the ith trial of a sequence of n independent trials. If X_i is 1 or 0 as the ith trial is a success or failure, then the problem remains invariant under permutations of the X's, and the UMP invariant test rejects (Example 6.3.6) when $\sum X_i > C$. It now follows from the remarks above that this test is also most stringent.

Another illustration is furnished by the general univariate linear hypothesis. Here it follows from the discussion in Example 8.5.2 that the standard test for testing $H : \eta_1 = \cdots = \eta_r = 0$ or $H' : \sum_{i=1}^r \eta_i^2/\sigma^2 \le \psi_0^2$ is most stringent.

When the invariance approach is not applicable, the explicit determination of most stringent tests typically is difficult. The following is a class of problems for which they are easily obtained by a direct approach. Let the distributions of X constitute a one-parameter exponential family, the density of which is given by (3.19), and consider the hypothesis $H : \theta = \theta_0$. Then according as $\theta > \theta_0$ or $\theta < \theta_0$, the envelope power $\beta^*_\alpha(\theta)$ is the power of the UMP one-sided test for testing H against $\theta > \theta_0$ or $\theta < \theta_0$. Suppose that there exists a two-sided test φ_0 given by (4.3), such that

$$\sup_{\theta<\theta_0}[\beta^*_\alpha(\theta) - \beta_{\varphi_0}(\theta)] = \sup_{\theta>\theta_0}[\beta^*_\alpha(\theta) - \beta_{\varphi_0}(\theta)], \tag{8.37}$$

and that the supremum is attained on both sides, say at points $\theta_1 < \theta_0 < \theta_2$. If $\beta_{\varphi_0}(\theta_i) = \beta_i, i = 1, 2$, an application of the fundamental lemma [Theorem 3.6.1(iii)] to the three points θ_1, θ_2, θ_0 shows that among all tests φ with $\beta_\varphi(\theta_1) \ge \beta_1$ and $\beta_\varphi(\theta_2) \ge \beta_2$, only φ_0 satisfies $\beta_\varphi(\theta_0) \le \alpha$. For any other level-$\alpha$ test, therefore, either $\beta_\varphi(\theta_1) < \beta_1$ or $\beta_\varphi(\theta_2) < \beta_2$, and it follows that φ_0 is the unique most stringent test. The existence of a test satisfying (8.37) can be proved by a continuity consideration [with respect to variation of the constants C_i and γ_i which define the boundary of the test (4.3)] from the fact that for the UMP one-sided test against the alternatives $\theta > \theta_0$ the right-hand side of (8.37) is zero and the left-hand side positive, while the situation is reversed for the other one-sided test.

8.7 Monotone Tests

In some testing problems, it may be reasonable to restrict attention to tests that are monotone in an appropriate sense. We begin with a motivating example.

Example 8.7.1 (**Testing For Superiority**) Suppose $(X_1, \dots, X_s)^\top$ is multivariate normal with unknown mean $\theta = (\theta_1, \dots, \theta_s)^\top$ and known covariance matrix Σ. Assume $\Sigma = I_s$, the identity. The null hypothesis $H_0 : \theta \in \Omega_0$ specfies not all $\theta_i > 0$ so that

$$\Omega_0 = \{\theta : \ \theta_i \le 0 \quad \text{for some } i\} \ ,$$

and the alternative specifies all $\theta_i > 0$, or $\Omega_1 = \Omega_0^c$. By Problem 4.8, the only unbiased test is $\phi \equiv \alpha$, and so trivially it is UMPU. However, there are reasonable tests for this problem. The likelihood ratio test rejects for large values of $T = \min(X_1, \ldots, X_s)$ (Problem 8.41), say when T exceeds an appropriate threshold c. In order to determine the critical value c so that the level of the test is controlled, we must ensure that

$$P_\theta\{T > c\} \le \alpha \ , \quad \text{for all } \theta \in \Omega_0 \ .$$

As a function of θ, this rejection probability is increasing in each of the θ_i. Hence, this probability is maximized over $\theta \in \Omega_0$ when one of the θ_i is zero and the remaining are tending to ∞. By symmetry, let $\theta_1 = 0$. Then, for $\theta \in \Omega_0$,

$$P_\theta\{T > c\} \le P_{\theta_1=0}\{X_1 > c\}$$

and

$$\sup_{\theta \in \Omega_0} P_\theta\{T > c\} = P_{\theta_1=0}\{X_1 > c\} \ ,$$

because for any c and $i > 1$, $P_{\theta_i}\{X_i > c\} \to 0$ as $\theta_i \to \infty$. Hence, the test that rejects when $T > z_{1-\alpha}$ has size α. Such a test seems intuitively reasonable, for in order to claim that all θ_i are positive, large values of the smallest X_i support this claim. However, we would like to know if there is a more principled reason to support this test. Note that when all $\theta_i = 0$, the rejection probability becomes

$$P_0\{\min(X_1, \ldots, X_s) > z_{1-\alpha}\} = \alpha^s \ ,$$

which is $< \alpha$. Hence, the likelihood ratio test is biased (since the power near the origin will be $< \alpha$). Also notice that the likelihood ratio test is monotone in the following sense. If $x = (x_1, \ldots, x_s)^\top$ is in the rejection region and $x_i' \ge x_i$ for all i, then $x' = (x_1', \ldots, x_s')^\top$ is also in the rejection region. Intuitively, if you are willing to reject based on X, you should be willing to reject based on X' if all the components of X' are at least as large as the corresponding components of X. We will return to this example shortly. ■

The principle of monotonicity is based on the following idea. Suppose $(X_1, \ldots, X_s)^\top$ is distributed according to P_θ, where $\theta = (\theta_1, \ldots, \theta_s)^\top$. Let $\theta' = (\theta_1', \ldots, \theta_s')^\top$. If $\theta \le \theta'$ (meaning $\theta_i \le \theta_i'$ for all i), then we will assume that θ' typically produces larger values of X than does θ. In order to formalize the sense in which the P_θ are appropriately stochastically increasing, we will generalize the univariate definition in Section 3.4 to s dimensions.

A set $\omega \in \mathbf{R}^s$ is said to be monotone increasing if

$$x = (x_1, \ldots, x_s)^\top \in \omega \ \text{ and } x_i \le x_i' \text{ for all } i \text{ implies } x' \in \omega \ . \tag{8.38}$$

(Similarly, ω is monotone decreasing if $-\omega = \{x : -x \in \omega\}$ is monotone increasing.) A nonrandomized test is called monotone increasing, or just monotone, if its rejection region is monotone increasing in the sense (8.38). In this section, we will restrict attention to nonrandomized tests. (More generally, one may call a test $\phi = \phi(X_1, \ldots, X_n)$ monotone if it is nondecreasing in each of its arguments.) The restriction to monotone rejection regions may be appropriate when the distributions P_θ are stochastically increasing; that is, $\theta_i \le \theta_i'$ for all i implies

$$\int_\omega dP_\theta \le \int_\omega dP_{\theta'} \tag{8.39}$$

for every monotone increasing set ω.

We now consider sufficient conditions to verify (8.39). Let $\theta \le \theta'$ (componentwise) and $X \sim P_\theta$. Suppose we can find a function h such that $X' = h(X)$, with $X \le X'$ and $X' \sim P_{\theta'}$. Another possibility is to assume that, for some random vector Z, we may write $X = f(Z)$ and $X' = f'(Z)$, where f and f' are such that $X \le X'$, $X \sim P_\theta$, and $X' \sim P_{\theta'}$. Then (8.39) holds in either case because

$$\int_\omega dP_\theta = P\{X \in \omega\} \le P\{X' \in \omega\} = \int_\omega dP_{\theta'} \ .$$

For example, in a general multivariate location model, P_θ is the distribution of $Z + \theta$ where Z is a random vector with any fixed known distribution. If we take $f(Z) = Z + \theta$ and $f'(Z) = Z + \theta'$, then the conditions are satisfied. Thus, Condition (8.39) holds in Example 8.7.1.

Assume, as in Example 8.7.1, that the alternative parameter space Ω_1 is monotone increasing. Then, the case for the restriction to monotone tests is particularly compelling when, as in the above sufficient conditions for (8.39), we can write $X = f(Z)$, $X' = f'(Z)$ and $X \le X'$, where $X \sim P_\theta$ and $X' \sim P_{\theta'}$ with $\theta \le \theta'$. Consider the following (unrealistic) situation. Michael is testing H_0 on the basis of $X = f(Z)$ and Azeem is testing H_0 on the basis of $X' = f'(Z)$ (for the same Z, so that $X \le X'$). Suppose Michael rejects H_0 and claims $\theta \in \Omega_1$. Michael would also conclude $\theta' \in \Omega_1$ for any $\theta' \ge \theta$, and so Azeem should reject H_0 as well. Thus, if Michael and Azeem are using the same test ϕ, then when Michael rejects H_0 on the basis of X, so must Azeem on the basis of X'. Since $X \le X'$, the test must be monotone. Finally, even if Michael's and Azeem's testing problems are not coupled, Michael should still want to apply the same test ϕ when faced with the problem of testing H_0 based on X.

The restriction to monotone tests sometimes allows one to find an optimal procedure, as we now show.

Example 8.7.2 (**Continuation of Example** 8.7.1) We will show that, among monotone level α tests, the likelihood ratio test is UMP. Recall that the rejection region is the set

$$E = \{(x_1, \ldots, x_s)^\top : \ \min(x_1, \ldots, x_s) \ge z_{1-\alpha}\} \ .$$

Suppose E' is any other monotone increasing rejection region and that E' includes some point y that is not in E. Therefore, $y_i < z_{1-\alpha}$ for some i. Without loss of generality, assume $y_1 < z_{1-\alpha}$. Then, since E' is monotone, $y \in E'$ implies the set $R_y \subset E'$, where

$$R_y = \{(x_1, \ldots, x_s)^\top : x_i \geq y_i \quad \text{for all } i\} .$$

Then, the size of the test with rejection region E' is at least

$$\sup_{\theta \in \Omega_0} P_\theta\{R_y\} \geq \sup_{\theta_1 = 0, \theta_i \to \infty} P_\theta\{X_i \geq y_i,\ i = 1, \ldots, s\}$$

$$= P_0\{X_1 \geq y_1\} > P_0\{X_1 > z_{1-\alpha}\} = \alpha .$$

Hence, such a test is not level α. Therefore, any other level α test with rejection region E' must satisfy $E' \subseteq E$. If E' is a strict subset of E with the set difference $E \setminus E'$ having positive Lebesgue measure, then the power against any alternative of the test with rejection region E is strictly bigger than that of E', as claimed.

Despite the likelihood ratio test being UMP among monotone tests, it is not α-admissible. In order to construct a level α test with better power, one can enlarge the rejection region E of the likelihood ratio test in such a way so that the new rejection region is still level α. Since the rejection probability of the likelihood ratio test is α^s at the origin, one can add a region F to E where F is some region in the quadrant where all components are negative. For example, one may consider

$$F = \{x : \max_i x_i \leq d\}$$

for some $d < 0$. If d is chosen far enough away from 0 so that

$$P_0\{F\} < \alpha - \alpha^s ,$$

then the test that has rejection region $E \cup F$ has probability of a Type 1 error under $\theta = 0$ equal to

$$P_0\{E \cup F\} = P_0\{E\} + P_0\{F\} = \alpha^s + P_0\{F\} < \alpha .$$

One just needs to make sure that the level is controlled for all $\theta \in \Omega_0$ (Problem 8.43). Such a test is clearly nonsensical because $X \in F$ means that all components of X are negative, which is not evidence against the null hypothesis. While the test with rejection region $E \cup F$ increases power, it does so at the expense of increasing Type 1 error, i.e., the rejection probability under $\theta \in \Omega_0$. Although the likelihood ratio test is not α-admissible, it is in fact d-admissible; see Cohen et al. (1983) and Nomakuchi and Sakata (1987). Perlman and Wu (1999) advocate for the use of the likelihood ratio test and question the utility of the notion of α-admissibility. ■

Monotonicity does not always reduce the problem sufficiently far so that a UMP monotone test exists. However, as the next example shows, we can obtain the maximin monotone level α test, i.e., the test that maximizes minimum power among monotone level α tests.

Example 8.7.3 (**Moment Inequalities**) Assume $X = (X_1, \dots, X_s)^\top$ is multivariate normal with unknown mean $\theta = (\theta_1, \dots, \theta_s)^\top$ and known invertible covariance matrix Σ. The problem now is to test the null hypothesis $H_0 : \theta \in \Omega_0$, where

$$\Omega_0 = \{\theta : \ \theta_i \le 0 \ \text{ for all } i = 1, \dots s\} .$$

Note that $\Omega_1 = \Omega_0^c$ is a monotone increasing set. As in Example 8.7.1, the UMPU level α test is the trivial test $\phi \equiv \alpha$, but monotonicity considerations apply.[4]

First, let's consider the likelihood ratio test. It rejects for large values of T given by

$$T = \inf_{\theta \in \Omega_0} (X - \theta)^\top \Sigma^{-1} (X - \theta) . \tag{8.40}$$

One can check that T is monotone increasing in each of its arguments, and therefore the test that rejects when $T > c$ is monotone. For this problem, any monotone test has its largest probability of rejection under $\theta \in \Omega_0$ occur at $\theta = 0$. Therefore, the constant $c = c_{1-\alpha}$ should be determined so that

$$P_0\{T \ge c_{1-\alpha}\} = \alpha .$$

In the special case that Σ is the identity matrix, the test statistic T reduces to $T = \sum_i \max^2(X_i, 0)$.

In order to determine the maximin monotone test, let us assume that Σ has diagonal elements equal to one (or simply divide each X_i by its standard deviation). Consider the region in the alternative parameter space $\omega(\epsilon)$ defined, for some fixed $\epsilon > 0$, by

$$\omega(\epsilon) = \{\theta : \ \theta_i \ge \epsilon \quad \text{for some } i\} .$$

The goal is to maximize

$$\inf_{\theta \in \omega(\epsilon)} P_\theta\{\text{reject } H_0\}$$

among level α monotone tests. Letting $M = \max(X_1, \dots, X_s)$, consider the test that rejects H_0 if $M > d$ for some constant d. Such a test is clearly monotone and $d = d_{1-\alpha}$ can be determined so that

$$P_0\{M \ge d_{1-\alpha}\} = \alpha .$$

[4] When the null hypothesis parameter space is described as a number of inequalities about means being satisfied, the problem is known in econometrics as testing moment inequalities; for a review, see Canay and Shaikh (2017).

Of course, if Σ is the identity matrix, then

$$d_{1-\alpha} = \Phi^{-1}\left[(1-\alpha)^{1/s}\right] . \tag{8.41}$$

We now argue that the test that rejects when $M \geq d_{1-\alpha}$ is the maximin monotone test. The worst-case power of this test over the region $\omega(\epsilon)$ is

$$\inf_{\theta \in \omega(\epsilon)} P_\theta\{M \geq d_{1-\alpha}\} ,$$

which occurs when some $\theta_i = \epsilon$ and the remaining θ_j satisfy $\theta_j \to -\infty$. In such a case, $M = X_i$ with probability tending to one, and the worse case power becomes

$$P_{\theta_i=\epsilon}\{X_i \geq d_{1-\alpha}\} = 1 - \Phi(d_{1-\alpha} - \epsilon) . \tag{8.42}$$

Assume ϕ is another monotone level α test with acceptance region A and rejection region $R = A^c$. Suppose ϕ includes some point $x \in A$ that falls in the interior of the rejection region of the test based on M, so that $x_i > d_{1-\alpha}$ for some i. Without loss of generality, assume $x_1 > d_{1-\alpha}$. Let

$$A_x = \{y : y_i \leq x_i \quad \text{for all } i\} .$$

Then, A_x is a monotone decreasing set and $x \in A$ imples $A_x \subseteq A$. Therefore, $P_\theta\{A\} \geq P_\theta\{A_x\}$ and so

$$\sup_{\theta \in \omega(\epsilon)} P_\theta\{A\} \geq \sup_{\theta \in \omega(\epsilon)} P_\theta\{X_1 \leq x_1, \ldots, X_s \leq x_s\} .$$

But as $\theta_j \to -\infty$, the event $\{X_j \leq x_j\}$ has probability tending to one. So,

$$\sup_{\theta \in \omega(\epsilon)} P_\theta\{A\} \geq \lim_{\theta_1=\epsilon, \theta_j \to -\infty, j>1} P_\theta\{X_1 \leq x_1, \ldots, X_s \leq x_s\} =$$

$$P_{\theta_1=\epsilon}\{X_1 \leq x_1\} = \Phi(x_1 - \epsilon) .$$

Therefore,

$$\inf_{\theta \in \omega(\epsilon)} P_\theta(R) \leq 1 - \Phi(x_1 - \epsilon) < 1 - \Phi(d_{1-\alpha} - \epsilon) ,$$

since $x_1 > d_{1-\alpha}$. Therefore, the worst-case power of ϕ is worse than that of the claimed optimal test, by (8.42). The assumption that $x_1 > d_{1-\alpha}$ cannot hold in order for ϕ to be optimal. Hence, the rejection region R of ϕ must be contained in the rejection region $\{M \geq d_{1-\alpha}\}$. But then it cannot have better power than the test based on M. Hence, the test that rejects for large M is maximin among monotone tests. Note that, by taking $\epsilon = 0$ in the argument, the test is also the monotone test with smallest worst-case bias; that is,

$$\alpha - \inf_{\theta \in \Omega_0^c} P_\theta\{\text{reject H}_0\}$$

is minimized among monotone tests.

In this example, the case for tests that are not monotone is motivated by the following observations. First, the critical value of the test based on M increases with the dimension s; see (8.41) in the case where Σ is the identity. Therefore, one strategy is to reduce the dimension so that a reduced critical value makes it easier to reject H_0. For example, if one observes a large negative X_j, one might feel confident that $\theta_j \leq 0$ and remove such an index j from consideration. That is, for some threshold $t < 0$, let $J = \{j : X_j > t\}$, and only test θ_j for $j \in J$. For example, one might use the statistic $\max_{j \in J} X_j$ with critical value determined when all corresponding $\theta_j = 0$. However, it is not clear that such a procedure controls the size of the test. Such procedures are known in the econometrics literature as moment selection procedures. Note that such a procedure is not monotone. As such, the probability of a Type 1 error as a function of $\theta \in \Omega$ need not be maximized at the origin, and one must take into account the selection step. Moment selection methods that provide error control are provided in Andrews and Barwick (2012) and Romano et al. (2014), who also provide references to this growing literature. ■

Unfortunately, the test that is maximin among monotone tests need not be maximin without the monotonicity restriction. Such is the case in Example 8.7.3, but we next show it in another example.

Example 8.7.4 (**Cauchy Location Model**) Let X be an observation from the Cauchy location model with center θ; X has density $f(x - \theta)$, where

$$f(x) = \frac{1}{\pi(1 + x^2)} \ .$$

Consider the problem of testing $H_0 : \theta = 0$ against $\theta > 0$. By Problem 3.33, no UMP test exists. The only monotone tests are those with rejection regions $\{X \geq c\}$ (or $\{X > c\}$) for some c. Hence, if $c = c_{1-\alpha}$ is the $1 - \alpha$ quantile of the Cauchy distributed centered at 0, then the test that rejects if $X \geq c_{1-\alpha}$ is trivially UMP among monotone tests.

Next, consider the problem of finding a maximin test over the region $\theta \geq a$ for some fixed $a > 0$, without the restriction to monotone tests. We now argue that the UMP monotone test is generally not maximin, depending on the value of a. First, note that the power of the UMP monotone test over $\theta \geq a$ attains its minimum at $\theta = a$. Also, for testing $\theta = 0$ against $\theta = a$, the likelihood ratio is

$$\frac{f(X - a)}{f(X)} = \frac{1 + X^2}{1 + (X - a)^2} \ , \tag{8.43}$$

which is ≥ 1 if and only if $X \geq a/2$. Therefore, by the Neyman–Pearson Lemma, the test that rejects if $X \geq a/2$ is most powerful at level $P_0\{X \geq a/2\}$. If $a/2 = c_{1-\alpha}$,

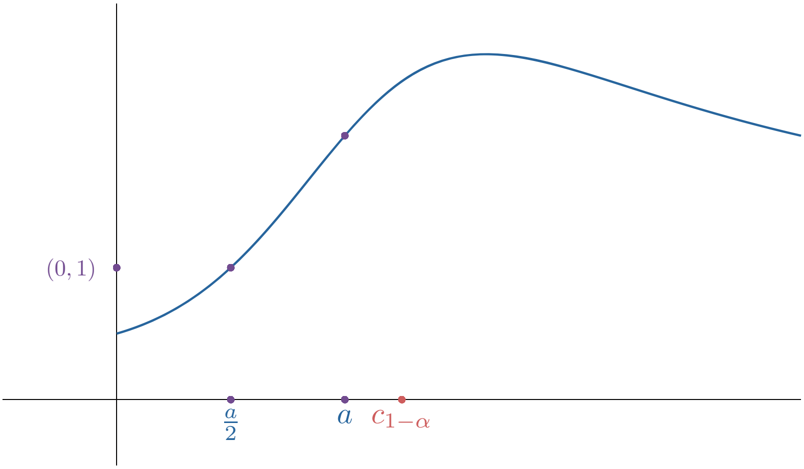

Figure 8.1 Cauchy likelihood ratio

then this MP level α test is in fact the UMP monotone level α test. It follows easily that, in this case, the test that rejects if $X \geq a/2$ is maximin among all level α tests. Indeed, if it were not, there would be another test whose power at $\theta = a$ would be less than that of the UMP monotone test, a contradiction since the UMP monotone test maximizes power at a.

On the other hand, consider the case where $a/2 < c_{1-\alpha}$. Then, the UMP monotone test is no longer maximin. To appreciate why, again consider the likelihood ratio for testing $\theta = 0$ against $\theta = a$. The likelihood ratio at $a/2$ is 1, it exceeds 1 for $X > a/2$ and it tends to 1 as $X \to \infty$. The likelihood ratio (8.43) is plotted in Figure 8.1 as a function of X. Since $a/2 < c_{1-\alpha}$, we may therefore modify the rejection region $\{X \geq c_{1-\alpha}\}$ by removing a small interval of very large X values and including values of X near $c_{1-\alpha}$. The new rejection region now includes an interval $[c_{1-\alpha} - \epsilon, c_{1-\alpha})$ for some small enough ϵ. The result is an increase in power at $\theta = a$, and it can be done in such a way that the worst-case power is not decreased (Problem 8.48).

By a similar argument, it follows that, again in the case $a/2 < c_{1-\alpha}$, the rejection region of the maximin test cannot include any semi-infinite interval (b, ∞). On the other hand, the rejection region of the maximin test cannot be a bounded set, because the probability of any bounded set tends to 0 as $\theta \to \infty$. It follows that the rejection region of the maximin test is unbounded but does not include a semi-infinite interval. That is, to the right of any point b, both the rejection region and the acceptance region contains subsets with positive measure. A conjecture is that it is an infinite sequence of intervals, though it is clearly not monotone.

Finally, in the case $a/2 > c_{1-\alpha}$, similar arguments show that the maximin monotone level α test is not maximin among all level α tests. ■

8.8 Problems

Section 8.1

Problem 8.1 *Existence of maximin tests.*[5] Let $(\mathcal{X}, \mathcal{A})$ be a Euclidean sample space, and let the distributions $P_\theta, \theta \in \Omega$, be dominated by a σ-finite measure over $(\mathcal{X}, \mathcal{A})$. For any mutually exclusive subsets Ω_H, Ω_K of Ω there exists a level-α test maximizing (8.2).
[Let $\beta = \sup[\inf_{\Omega_k} E_\theta\varphi(X)]$, where the supremum is taken over all level-α tests of $H : \theta \in \Omega_H$. Let φ_n be a sequence of level-α tests such that $\inf_{\Omega_K} E_\theta\varphi_n(X)$ tends to β. If φ_{n_i} is a subsequence and φ a test (guaranteed by Theorem A.5.1 of the Appendix) such that $E_\theta\varphi_{n_i}(X)$ tends to $E_\theta\varphi(X)$ for all $\theta \in \Omega$, then φ is a level-α test and $\inf_{\Omega_k} E_\theta\varphi(X) = \beta$.]

Problem 8.2 In Example 8.1.1, explain why the maximin test is not UMPU for the alternatives considered.

Problem 8.3 In Example 8.1.3, complete the argument using Corollary 8.1.1 to find the maximin test without assuming you already know the UMPI test. What if the alternative specifies $\sum_{i=1}^n \xi_i^2 \geq \delta^2$?

Problem 8.4 *Locally most powerful tests.* [6] Let d be a measure of the distance of an alternative θ from a given hypothesis H. A level-α test φ_0 is said to be *locally most powerful* (LMP) if, given any other level-α test φ, there exists Δ such that

$$\beta_{\varphi_0}(\theta) \geq \beta_\varphi(\theta) \quad \text{for all } \theta \text{ with } 0 < d(\theta) < \Delta. \tag{8.44}$$

Suppose that θ is real-valued and that the power function of every test is continuously differentiable at θ_0.

(i) If there exists a unique level-α test φ_0 of $H : \theta = \theta_0$, maximizing $\beta'_\varphi(\theta_0)$, then φ_0 is the unique LMP level-α test of H against $\theta > \theta_0$ for $d(\theta) = \theta - \theta_0$.

(ii) To see that (i) is not correct without the uniqueness assumption, let X take on the values 0 and 1 with probabilities $P_\theta(0) = \frac{1}{2} - \theta^3$, $P_\theta(1) = \frac{1}{2} + \theta^3$, $-\frac{1}{2} < \theta^3 < \frac{1}{2}$, and consider testing $H : \theta = 0$ against $K : \theta > 0$. Then every test φ of size α maximizes $\beta'_\varphi(0)$, but not every such test is LMP. [Kallenberg et al. (1984).]

(iii) The following[7] is another counterexample to (i) without uniqueness, in which in fact no LMP test exists. Let X take on the values 0, 1, 2 with probabilities

[5] The existence of maximin tests is established in considerable generality in Cvitanic and Karatzas Karatzas (2001).

[6] Locally optimal tests for multiparameter hypotheses are given in Gupta and Vermeire (1986).

[7] Due to John Pratt.

$$P_\theta(x) = \alpha + \epsilon\left[\theta + \theta^2 \sin\left(\frac{x}{\theta}\right)\right] \qquad \text{for} \quad x = 1, 2,$$
$$P_\theta(0) = 1 - p_\theta(1) - p_\theta(2),$$

where $-1 \le \theta \le 1$ and ϵ is a sufficiently small number. Then a test φ at level α maximizes $\beta'(0)$ provided

$$\varphi(1) + \varphi(2) = 1\ ,$$

but no LMP test exists.

(iv) A unique LMP test maximizes the minimum power locally provided its power function is bounded away from α for every set of alternatives which is bounded away from H.

(v) Let $X_1, \ldots, X_n$ be a sample from a Cauchy distribution with unknown location parameter θ, so that the joint density of the X's is $\pi^{-n}\prod_{i=1}^n[1 + (x_i - \theta)^2]^{-1}$. The LMP test for testing $\theta = 0$ against $\theta > 0$ at level $\alpha < \frac{1}{2}$ is not unbiased and hence does not maximize the minimum power locally.

[(iii): The unique most powerful test against θ is

$$\begin{cases} \varphi(1) \\ \varphi(2) \end{cases} = 1 \quad \text{if } \sin\left(\frac{1}{\theta}\right) \gtreqless \sin\left(\frac{2}{\theta}\right),$$

and each of these inequalities holds at values of θ arbitrarily close to 0.

(v): There exists M so large that any point with $x_i \ge M$ for all $i = 1, \ldots, n$ lies in the acceptance region of the LMP test. Hence the power of the test tends to zero as θ tends to infinity.]

Problem 8.5 Under the setting of Problem 3.35, determine the locally most powerful test.

Problem 8.6 A level-α test φ_0 is locally unbiased (loc. unb.) if there exists $\Delta_0 > 0$ such that $\beta_{\varphi_0}(\theta) \ge \alpha$ for all θ with $0 < d(\theta) < \Delta_0$; it is LMP loc. unb. if it is loc. unb. and if, given any other loc. unb. level-α test φ, there exists Δ such that (8.44) holds. Suppose that θ is real-valued and that $d(\theta) = |\theta - \theta_0|$, and that the power function of every test is twice continuously differentiable at $\theta = \theta_0$.

(i) If there exists a unique test φ_0 of $H : \theta = \theta_0$ against $K : \theta \ne \theta_0$ which among all loc. unb. tests maximizes $\beta''(\theta_0)$, then φ_0 is the unique LMP loc. unb. level-α test of H against K.

(ii) The test of part (i) maximizes the minimum power locally provided its power function is bounded away from α for every set of alternatives that is bounded away from H.

[(ii): A necessary condition for a test to be locally minimax is that it is loc. unb.]

Problem 8.7 *Locally uniformly most powerful tests.* If the sample space is finite and independent of θ, the test φ_0 of Problem 8.4(i) is not only LMP but also locally uniformly most powerful (LUMP) in the sense that there exists a value $\Delta > 0$ such that φ_0 maximizes $\beta_\varphi(\theta)$ for all θ with $0 < \theta - \theta_0 < \Delta$.

Problem 8.8 The following two examples show that the assumption of a finite-sample space is needed in Problem 8.7.

(i) Let $X_1, \ldots, X_n$ be i.i.d. according to a normal distribution $N(\sigma, \sigma^2)$ and test $H : \sigma = \sigma_0$ against $K : \sigma > \sigma_0$.
(ii) Let X and Y be independent Poisson variables with $E(X) = \lambda$ and $E(Y) = \lambda + 1$, and test $H : \lambda = \lambda_0$ against $K : \lambda > \lambda_0$. In each case, determine the LMP test and show that it is not LUMP.

[Compare the LMP test with the most powerful test against a simple alternative.]

Section 8.2

Problem 8.9 Let the distribution of X depend on the parameters $(\theta, \vartheta) = (\theta_1, \ldots, \theta_r, \vartheta_1, \ldots, \vartheta_s)$. A test of $H : \theta = \theta^0$ is *locally strictly unbiased* if for each φ, (a) $\beta_\varphi(\theta^0, \varphi) = \alpha$, (b) there exists a θ-neighborhood of θ^0 in which $\beta_\varphi(\theta, \vartheta) > \alpha$ for $\theta \neq \theta^0$.

(i) Suppose that the first and second derivatives

$$\beta^i_\varphi(\vartheta) = \frac{\partial}{\partial\theta_i}\beta_\varphi(\theta, \vartheta)\bigg|_{\theta^0} \quad \text{and} \quad \beta^{ij}_\varphi(\vartheta) = \frac{\partial^2}{\partial\theta_i\partial\theta_j}\beta_\varphi(\theta, \vartheta)\bigg|_{\theta^0}$$

exist for all critical functions φ and all ϑ. Then a necessary and sufficient condition for φ to be locally strictly unbiased is that $\beta'_\varphi = 0$ for all i and ϑ, and that the matrix $(\beta^{ij}_\varphi(\vartheta))$ is positive definite for all ϑ.
(ii) A test of H is said to be of *type E* (*type D* is $s = 0$ so that there are no nuisance parameters) if it is locally strictly unbiased and among all tests with this property maximizes the determinant $|(\beta^{ij}_\varphi)|$.[8] (This determinant under the stated conditions turns out to be equal to the Gaussian curvature of the power surface at θ^0.) Then the test φ_0 given by (7.7) for testing the general linear univariate hypothesis (7.3) is of type E.

[(ii): With $\theta = (\eta_1, \ldots, \eta_r)$ and $\vartheta = (\eta_{r+1}, \ldots, n_s, \sigma)$, the test φ_0, by Problem 7.5, has the property of maximizing the surface integral

$$\int_S [\beta_\varphi(\eta, \sigma^2) - \alpha]\, dA$$

[8] An interesting example of a type-D test is provided by Cohen and Sackrowitz (1975), who show that the χ^2-test of Chapter 16.3 has this property. Type D and E tests were introduced by Isaacson (1951).

among all similar (and hence all locally unbiased) tests where $S = \{(\eta_1, \dots, \eta_r) : \sum_{i=1}^r \eta_i^2 = \rho^2\sigma^2\}$. Letting ρ tend to zero and utilizing the conditions

$$\beta_\varphi^i(\vartheta) = 0, \qquad \int_S \eta_i \eta_j \, dA = 0 \quad \text{for } i \neq j, \qquad \int_S \eta_i^2 \, dA = k(\rho\sigma),$$

one finds that φ_0 maximizes $\sum_{i=1}^r \beta_\varphi^{ii}(\eta, \sigma^2)$ among all locally unbiased tests. Since for any positive definite matrix, $|(\beta_\varphi^{ij})| \leq \prod \beta_\varphi^{ii}$, it follows that for any locally strictly unbiased test φ,

$$|(\beta_\varphi^{ij})| \leq \prod \beta_\varphi^{ii} \leq \left[\frac{\Sigma\beta_\varphi^{ii}}{r}\right]^r \leq \left[\frac{\Sigma\beta_{\varphi_0}^{ii}}{r}\right]^r = [\beta_{\varphi_0}^{11}]^r = |(\beta_{\varphi_0}^{ij})|.]$$

Problem 8.10 Let $Z_1, \dots, Z_n$ be identically independently distributed according to a continuous distribution D, of which it is assumed only that it is symmetric about some (unknown) point. For testing the hypothesis $H : D(0) = \frac{1}{2}$, the sign test maximizes the minimum power against the alternatives $K : D(0) \leq q (q < \frac{1}{2})$. [A pair of least favorable distributions assign probability 1, respectively, to the distributions $F \in H$, $G \in K$ with densities

$$f(x) = \frac{1-2q}{2(1-q)} \left(\frac{q}{1-q}\right)^{[|x|]}, \qquad g(x) = (1-2q)\left(\frac{q}{1-q}\right)^{|[x]|}$$

where for all x (positive, negative, or zero) $[x]$ denotes the largest integer $\leq x$.]

Problem 8.11 Let $f_\theta(x) = \theta g(x) + (1-\theta)h(x)$ with $0 \leq \theta \leq 1$. Then $f_\theta(x)$ satisfies the assumptions of Lemma 8.2.1 provided $g(x)/h(x)$ is a nondecreasing function of x.

Problem 8.12 Let $x = (x_1, \dots, x_n)$, and let $g_\theta(x, \xi)$ be a family of probability densities depending on $\theta = (\theta_1, \dots, \theta_r)$ and the real parameter ξ, and jointly measurable in x and ξ. For each θ, let $h_\theta(\xi)$ be a probability density with respect to a σ-finite measure ν such that $p_\theta(x) = \int g_\theta(x, \xi) h_\theta(\xi) \, d\nu(\xi)$ exists. We shall say that a function f of two arguments $u = (u_1, \dots, u_r)$, $v = (v_1, \dots, v_s)$ is nondecreasing in (u, v) if $f(u', v)/f(u, v) \leq f(u', v')/f(u, v')$ for all (u, v) satisfying $u_i \leq u_i'$, $v_j \leq v_j'$ $(i = 1, \dots, r; j = 1, \dots, s)$. Then $p_\theta(x)$ is nondecreasing in (x, θ) provided the product $g_\theta(x, \xi)h_\theta(\xi)$ is (a) nondecreasing in (x, θ) for each fixed ξ; (b) nondecreasing in (θ, ξ) for each fixed x; (c) nondecreasing in (x, ξ) for each fixed θ.
[Interpreting $g_\theta(x, \xi)$ as the conditional density of x given ξ, and $h_\theta(\xi)$ as the a priori density of ξ, let $\rho(\xi)$ denote the a posteriori density of ξ given x, and let $\rho'(\xi)$ be defined analogously with θ' in place of θ. That $p_\theta(x)$ is nondecreasing in its two arguments is equivalent to

$$\int \frac{g_\theta(x', \xi)}{g_\theta(x, \xi)} \rho(\xi)\, d\nu(\xi) \le \int \frac{g_{\theta'}(x', \xi)}{g_{\theta'}(x, \xi)} \rho'(\xi)\, d\nu(\xi).$$

By (a) it is enough to prove that

$$D = \int \frac{g_\theta(x', \xi)}{g_\theta(x, \xi)} [\rho'(\xi) - \rho(\xi)]\, d\nu(\xi) \ge 0.$$

Let $S_- = \{\xi : \rho'(\xi)/\rho(\xi) < 1\}$ and $S_+ = \{\xi : \rho(\xi)/\rho(\xi) \ge 1\}$. By (b) the set S_- lies entirely to the left of S_+. It follows from (c) that there exists $a \le b$ such that

$$D = a \int_{S_-} [\rho'(\xi) - \rho(\xi)]\, d\nu(\xi) + b \int_{S_+} [\rho'(\xi) - \rho(\xi)]\, d\nu(\xi),$$

and hence that $D = (b - a) \int_{S_+} [\rho'(\xi) - \rho(\xi)]\, d\nu(\xi) \ge 0$.]

Problem 8.13 (i) Let X have binomial distribution $b(p, n)$, and consider testing $H : p = p_0$ at level α against the alternatives $\Omega_K : p/q \le \frac{1}{2} p_0/q_0$ or $\ge 2p_0/q_0$. For $\alpha = .05$ determine the smallest sample size for which there exists a test with power $\ge .8$ against Ω_K if $p_0 = .1, .2, .3, .4, .5$.

(ii) Let $X_1, \ldots, X_n$ be independently distributed as $N(\xi, \sigma^2)$. For testing $\sigma = 1$ at level $\alpha = .05$, determine the smallest sample size for which there exists a test with power $\ge .9$ against the alternatives $\sigma^2 \le \frac{1}{2}$ and $\sigma^2 \ge 2$.

[See Problem 4.5.]

Problem 8.14 *Double-exponential distribution.* Let $X_1, \ldots, X_n$ be a sample from the double-exponential distribution with density $\frac{1}{2} e^{-|x-\theta|}$. The LMP test for testing $\theta \le 0$ against $\theta > 0$ is the sign test, provided the level is of the form

$$\alpha = \frac{1}{2^n} \sum_{k=0}^{m} \binom{n}{k},$$

so that the level-α sign test is nonrandomized.

[Let R_k $(k = 0, \ldots, n)$ be the subset of the sample space in which k of the X's are positive and $n - k$ are negative. Let $0 \le k < l < n$, and let S_k, S_l be subsets of R_k, R_l such that $P_0(S_k) = P_0(S_l) \ne 0$. Then it follows from a consideration of $P_\theta(S_k)$ and $P_0(S_l)$ for small θ that there exists Δ such that $P_\theta(S_k) < P_\theta(S_l)$ for $0 < \theta < \Delta$. Suppose now that the rejection region of a nonrandomized test of $\theta = 0$ against $\theta > 0$ does not consist of the upper tail of a sign test. Then it can be converted into a sign test of the same size by a finite number of steps, each of which consists in replacing an S_k by an S_l with $k < l$, and each of which therefore increases the power for θ sufficiently small.]

Section 8.3

Problem 8.15 If (8.13) holds, show that q_1 defined by (8.11) belongs to $\mathcal{P}_1$.

Problem 8.16 Show that there exists a unique constant b for which q_0 defined by (8.11) is a probability density with respect to μ, that the resulting q_0 belongs to $\mathcal{P}_0$, and that $b \to \infty$ as $\epsilon_0 \to 0$.

Problem 8.17 Prove the formula (8.15).

Problem 8.18 Show that if $\mathcal{P}_0 \neq \mathcal{P}_1$ and ϵ_0, ϵ_1 are sufficiently small, then $Q_0 \neq Q_1$.

Problem 8.19 Evaluate the test (8.21) explicitly for the case that P_i is the normal distribution with mean ξ_i and known variance σ^2, and when $\epsilon_0 = \epsilon_1$.

Problem 8.20 Determine whether (8.21) remains the maximin test if in the model (8.20) G_i is replaced by G_{ij}.

Problem 8.21 Write out a formal proof of the maximin property outlined in the last paragraph of Section 8.3.

Section 8.4

Problem 8.22 Let $X_1, \ldots, X_n$ be independent and normally distributed with means $E(X_i) = \mu_i$ and variance 1. The test of $H : \mu_1 = \cdots = \mu_n = 0$ that maximizes the minimum power over $\omega' : \sum \mu_i \geq d$ rejects when $\sum X_i \geq C$.
[If the least favorable distribution assigns probability 1 to a single point, invariance under permutations suggests that this point will be $\mu_1 = \cdots = \mu_n = d/n$].

Problem 8.23[9] (i) In the preceding problem determine the maximin test if ω' is replaced by $\sum a_i \mu_i \geq d$, where the a's are given positive constants.
(ii) Solve part (i) with $Var(X_i) = 1$ replaced by $Var(X_i) = \sigma_i^2$ (known).
[(i): Determine the point $(\mu_1^*, \ldots, \mu_n^*)$ in ω' for which the MP test of H against $K : (\mu_1^*, \ldots, \mu_n^*)$ has the smallest power, and show that the MP test of H against K is a maximin solution.]

Problem 8.24 Let $X_1, \ldots, X_n$ be independent normal variables with variance 1 and means $\xi_1, \ldots, \xi_n$, and consider the problem of testing $H : \xi_1 = \cdots = \xi_n = 0$ against the alternatives $K = \{K_1, \ldots, K_n\}$, where $K_i : \xi_j = 0$ for $j \neq i, \xi_i = \xi$ (known and positive). Show that the problem remains invariant under permutation of the X's and that there exists a UMP invariant test ϕ_0 which rejects when $\sum e^{\xi X_i} > C$, by the following two methods.

(i) The order statistics $X_{(1)} < \cdots < X_{(n)}$ constitute a maximal invariant.

[9] Due to Fritz Scholz.

(ii) Let f_0 and f_i denote the densities under H and K_i respectively. Then the level-α test ϕ_0 of H versus $K' : f = (1/n)\sum f_i$ is UMP invariant for testing H versus K.

[(ii): If ϕ_0 is not UMP invariant for H versus K, there exists an invariant test ϕ_1 whose (constant) power against K exceeds that of ϕ_0. Then ϕ_1 is also more powerful against K'.]

Problem 8.25 The UMP invariant test ϕ_0 of Problem 8.24

(i) maximizes the minimum power over K;
(ii) is admissible.
(iii) For testing the hypothesis H of Problem 8.24 against the alternatives $K' = \{K_1, \ldots, K_n, K_1', \ldots, K_n'\}$, where under $K_i' : \xi_j = 0$ for all $j \neq i$, $\xi_i = -\xi$, determine the UMP test under a suitable group G', and show that it is both maximin and invariant.

[ii): Suppose ϕ' is uniformly at least as powerful as ϕ_0, and more powerful for at least one K_i, and let

$$\phi^*(x_1, \ldots, x_n) = \frac{\sum \phi'(x_{i_1}, \ldots, x_{i_n})}{n!},$$

where the summation extends over all permutations. Then ϕ^* is invariant, and its power is independent of i and exceeds that of ϕ_0.]

Problem 8.26 Suppose Problems 8.24–8.25 are modified so that the one nonzero mean may ξ or $-\xi$. How do the results change?

Problem 8.27 Suppose $X_1, \ldots, X_n$ are independent normal variables with $X_i \sim N(\xi_i, 1)$. The null hypothesis specifies all $\xi_i = 0$. Fix an integer $k \geq 1$. Suppose ω' specifies that at least k of the X_i have mean at least ξ, where ξ is known and positive. Determine a maximin test as explicitly as possible.

Problem 8.28 For testing $H : f_0$ against $K : \{f_1, \ldots, f_s\}$, suppose there exists a finite group $G = \{g_1, \ldots, g_N\}$ which leaves H and K invariant and which is transitive in the sense that given $f_j, f_{j'} (1 \leq j, j')$ there exists $g \in G$ such that $\bar{g} f_j = f_{j'}$. In generalization of Problems 8.24, 8.25, determine a UMP invariant test, and show that it is both maximin against K and admissible.

Problem 8.29 To generalize the results of the preceding problem to the testing of $H : f$ versus $K : \{f_\theta, \theta \in \omega\}$, assume:

(i) There exists a group G that leaves H and K invariant.
(ii) $\bar{G}$ is transitive over ω.
(iii) There exists a probability distribution Q over G which is right invariant in the sense of Section 8.4.

Determine a UMP invariant test, and show that it is both maximin against K and admissible.

Problem 8.30 Let $X_1, \ldots, X_n$ be independent normal with means $\theta_1, \ldots, \theta_n$ and variance 1.

(i) Apply the results of the preceding problem to the testing of $H : \theta_1 = \cdots = \theta_n = 0$ against $K : \sum \theta_i^2 = r^2$, for any fixed $r > 0$.
(ii) Show that the results of (i) remain valid if H and K are replaced by $H' : \sum \theta_i^2 \leq r_0^2$, $K' : \sum \theta_i^2 \geq r_1^2$ $(r_0 < r_1)$.

Problem 8.31 Suppose in Problem 8.30(i) the variance σ^2 is unknown and that the data consist of $X_1, \ldots, X_n$ together with an independent random variable S^2 for which S^2/σ^2 has a χ^2-distribution. If K is replaced by $\sum \theta_i^2/\sigma^2 = r^2$, then

(i) the confidence sets $\sum(\theta_i - X_i)^2/S^2 \leq C$ are uniformly most accurate equivariant under the group generated by the n-dimensional generalization of the group G_0 of Example 6.11.2, and the scale changes $X_i' = cX_i$, $S'^2 = c^2S^2$.
(ii) The confidence sets of (i) are minimax with respect to the measure μ given by

$$\mu[C(X, S^2)] = \frac{1}{\sigma^2}[\text{ volume of } C(X, S^2)].$$

[Use polar coordinates with $|\theta^2| = \sum \theta_i^2$.]

Section 8.5

Problem 8.32 Let $X = (X_1, \ldots, X_p)$ and $Y = (Y_1, \ldots, Y_p)$ be independently distributed according to p-variate normal distributions with zero means and covariance matrices $E(X_iX_j) = \sigma_{ij}$ and $E(Y_iY_j) = \Delta\sigma_{ij}$.

(i) The problem of testing $H : \Delta \leq \Delta_0$ remains invariant under the group G of transformations $X^* = XA$, $Y^* = YA$, where $A = (a_{ij})$ is any nonsingular $p \times p$ matrix with $a_{ij} = 0$ for $i > j$, and there exists a UMP invariant test under G with rejection region $Y_1^2/X_1^2 > C$.
(ii) The test with rejection region $Y_1^2/X_1^2 > C$ maximizes the minimum power for testing $\Delta \leq \Delta_0$ against $\Delta \geq \Delta_1$ $(\Delta_0 < \Delta_1)$.
[(ii): That the Hunt–Stein Theorem is applicable to G can be proved in steps by considering the group G_q of transformations $X_q' = \alpha_1 X_1 + \cdots + \alpha_q X_q$, $X_i' = X_i$ for $i = 1, \ldots, q-1, q+1, \ldots, p$, successively for $q = 1, \ldots, p-1$. Here $\alpha_q \neq 0$, since the matrix A is nonsingular if and only if $a_{ii} \neq 0$ for all i. The group product $(\gamma_1, \ldots, \gamma_q)$ of two such transformations $(\alpha_1, \ldots, \alpha_q)$ and $(\beta_1, \ldots, \beta_q)$ is given by $\gamma_1 = \alpha_q + \beta_1$, $\gamma_2 = a_2\beta_q + \beta_2$, $\ldots$, $\gamma_{q-1} = \alpha_{q-1}\beta_q + \beta_{q-1}$, $\gamma_q = \alpha_q, \beta_q$, which shows G_q to be isomorphic to a group of scale changes (multiplication of all components by β_q) and translations [addition of $(\beta_1, \ldots, \beta_{q-1}, 0)$].

The result now follows from the Hunt–Stein Theorem and Example 8.5.1, since the assumptions of the Hunt–Stein Theorem, except for the easily verifiable measurability conditions, concern only the abstract structure $(G, \mathcal{B})$, and not the specific realization of the elements of G as transformations of some space.]

Problem 8.33 Suppose that the problem of testing $\theta \in \Omega_H$ against $\theta \in \Omega_K$ remains invariant under G, that there exists a UMP almost invariant test φ_0 with respect to G, and that the assumptions of Theorem 8.5.1 hold. Then φ_0 maximizes $\inf_{\Omega_K}[w(\theta)E_\theta\varphi(X) + u(\theta)]$ for any weight functions $w(\theta) \geq 0, u(\theta)$ that are invariant under $\bar{G}$.

Problem 8.34 Suppose X has the multivariate normal distribution in $\mathbf{R}^k$ with unknown mean vector h and known positive definite covariance matrix C^{-1}. Consider testing $h = 0$ versus $|C^{1/2}h| \geq b$ for some $b > 0$, where $|\cdot|$ denotes the Euclidean norm.
(i) Show the test that rejects when $|C^{1/2}X|^2 > c_{k,1-\alpha}$ is maximin, where $c_{k,1-\alpha}$ denotes the $1-\alpha$ quantile of the Chi-squared distribution with k degrees of freedom.
(ii) Show that the maximin power of the above test is given $P\{\chi^2_k(b^2) > c_{k,1-\alpha}\}$, where $\chi^2_k(b^2)$ denotes a random variable that has the noncentral Chi-squared distribution with k degrees of freedom and noncentrality parameter b^2.

Problem 8.35 Suppose $X_1, \ldots, X_k$ are independent, with $X_i \sim N(\theta_i, 1)$. Consider testing the null hypothesis $\theta_1 = \cdots = \theta_k = 0$ against $\max|\theta_i| \geq \delta$, for some $\delta > 0$. Find a maximin level α test as explicitly as possible. Compare this test with the maximin test if the alternative parameter space were $\sum_i \theta_i^2 \geq \delta^2$. Argue they are quite similar for small δ. Specifically, consider the power of each test against $(\delta, 0, \ldots, 0)$ and show that it is equal to $\alpha + C_\alpha\delta^2 + o(\delta^2)$ as $\delta \to 0$, and the constant C_α is the same for both tests.

Section 8.6

Problem 8.36 *Existence of most stringent tests*. Under the assumptions of Problem 8.1 there exists a most stringent test for testing $\theta \in \Omega_H$ against $\theta \in \Omega - \Omega_H$.

Problem 8.37 Let $\{\Omega_\Delta\}$ be a class of mutually exclusive sets of alternatives such that the envelope power function is constant over each Ω_Δ and that $\cup\Omega_\Delta = \Omega - \Omega_H$, and let φ_Δ maximize the minimum power over Ω_Δ. If $\varphi_\Delta = \varphi$ is independent of Δ, then φ is most stringent for testing $\theta \in \Omega_H$.

Problem 8.38 Let $(Z_1, \ldots, Z_N) = (X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ be distributed according to the joint density (5.55), and consider the problem of testing $H : \eta = \xi$ against the alternatives that the X's and Y's are independently normally distributed with common variance σ^2 and means $\eta \neq \xi$. Then the permutation test with rejection region $|\bar{Y} - \bar{X}| > C[T(Z)]$, the two-sided version of the test (5.54), is most stringent.

[Apply Problem 8.37 with each of the sets Ω_Δ consisting of two points (ξ_1, η_1, σ), (ξ_2, η_2, σ) such that

$$\xi_1 = \zeta - \frac{n}{m+n}\delta, \qquad \eta_1 = \zeta + \frac{m}{m+n}\delta;$$
$$\xi_2 = \zeta + \frac{n}{m+n}\delta, \qquad \eta_2 = \zeta - \frac{m}{m+n}\delta$$

for some ζ and δ.]

Problem 8.39 Show that the UMP invariant test of Problem 8.24 is most stringent.

Section 8.7

Problem 8.40 Show that a region ω is monotone increasing if and only if its complement is monotone decreasing. In the plane, how would you characterize the class of all monotone increasing regions?

Problem 8.41 In Example 8.7.1, determine the likelihood ratio test for general Σ, and show that it reduces to the test that rejects for large values of $\min(X_1, \ldots, X_s)$ when Σ is the identity matrix. How do you calculate the critical value for general Σ? Is the resulting test monotone?

Problem 8.42 Suppose $X = (X_1, \ldots, X_s)^\top \sim P_\theta$, where the P_θ form a multivariate location model. So P_θ is the distribution of $Z + \theta$, where Z has a fixed (known) distribution in $\mathbb{R}$. For testing superiority as in Example 8.7.1, determine the UMP monotone level α test as explicitly as possible.

Problem 8.43 In Example 8.7.1, show how one may add a region F to the rejection region E of the likelihood ratio test and still maintain the size of the test.

Problem 8.44 Suppose $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$ with both parameters unknown. Show that, for testing $\mu \le 0$ against $\mu > 0$, the one-sided t-test is *not* monotone increasing. Does the assumption (8.39) hold for the parametrization $(\theta_1, \theta_2) = (\mu, \sigma)$ or perhaps $(\mu, 1/\sigma)$? [Consider the monotone sets $\{X_1 > c\}$ when c is both positive and negative.]

Problem 8.45 In Example 8.7.3, find the most powerful test for testing $\theta \in \Omega_0$ against a fixed alternative $\theta = a$ and compute the power of this test. [The least favorable distribution puts mass one at the point θ_a, where θ_a minimizes $(\theta - a)^\top \Sigma^{-1} (\theta - a)$ over θ.]

Problem 8.46 In Example 8.7.3, assume Σ is the identity matrix. Calculate the minimum power of the likelihood ratio test over the region $\omega(\epsilon)$ and compare it to the maximin monotone test.

Problem 8.47 Find the maximin monotone level α test in Example 8.7.3 for general Σ. Also allow the region $\omega(\epsilon)$ to be generalized and have the form $\{\theta : \theta_i \geq \epsilon_i \text{ for some } i\}$, where the ϵ_i may vary with i.

Problem 8.48 Provide the missing details in Example 8.7.4. What happens in the case $a > 2c_{1-\alpha}$?

8.9 Notes

The concepts and results of Section 8.1 are essentially contained in the minimax theory developed by Wald for general decision problems. An exposition of this theory and some of its applications is given in Wald's book (1950). For more recent assessments of the important role of the minimax approach, see Brown (1994, 2000). The ideas of Section 8.3, and in particular Theorem 8.3.1, are due to Huber (1965) and form the core of his theory of robust tests [Huber (1981, Chapter 10)]. The material of Sections 8.4 and 8.5, including Lemma 8.4.1, Theorem 8.5.1, and Example 8.5.2, constitutes the main part of an unpublished paper of Hunt and Stein (1946).

Section 8.7 was inspired by Lehmann (1952a). Problem 8.45 is taken from Romano, Shaikh, and Wolf (2014). The α-inadmissibility of the likelihood ratio test in Example 8.7.1 has been studied in Berger (1989), Liu and Berger (1995) and McDermott and Wang (2002). Perlman and Wu (1999) advocate the use of likehood ratio tests and reject the utility of α-admissibility; they provide numerous examples and references.